AMERICAN JOURNAL OF MATHEMATICS

FOUNDED BY THE JOHNS HOPKINS UNIVERSITY

PUBLISHED UNDER THE JOINT AUSPICES OF

THE JOHNS HOPKINS UNIVERSITY
AND
THE AMERICAN MATHEMATICAL SOCIETY

Volume LXIII, Number 3
JULY, 1941

THE JOHNS HOPKINS PRESS
BALTIMORE, MARYLAND
U. S. A.

CONTENTS

The American Journal of Mathematics will appear four times yearly.

The subscription price of the Journal for the current volume is $7.50 (foreign postage 50 cents); single numbers $2.00.

A few complete sets of the Journal remain on sale.

Papers intended for publication in the Journal may be sent to any of the Editors.

Editorial communications may be sent to Professor F. D. Murnaghan at The Johns Hopkins University.

Subscriptions to the Journal and all business communications should be sent to The Johns Hopkins Press, Baltimore, Maryland, U. S. A.

Entered as second-class matter at the Baltimore, Maryland, Postoffice, acceptance for mailing at special rate of postage provided for in Section 1103, Act of October 3, 1917, Authorized on July 3, 1918.

PRINTED IN THE UNITED STATES OF AMERICA
BY J. H. FURST COMPANY, BALTIMORE, MARYLAND

a
op
to
th
us
in
iso
res
gr
ap
iso

con
We
(al
res
sha
wh
res

last
alge
the
tion
It s
imp
alge

Jour

Univ

CLASSES OF RESTRICTED LIE ALGEBRAS OF CHARACTERISTIC p. I *

By N. Jacobson.

0. 1. In a previous paper [1] we discussed the Lie algebras, obtained from a simple associative algebra of characteristic 0 by defining the Lie algebra operation $[a, b]$ as $ab - ba$, and the subalgebras of skew elements of involutorial associative algebras of this type. The present paper develops a similar theory for algebras of characteristic p. The fundamental concept that enables us to obtain our extension is that of a restricted Lie algebra. The operations in these systems are $a + b, a\alpha, \alpha$ in the field, $[a, b]$ and a^p. Conditions for isomorphisms, the derivation algebras and the automorphism groups for the restricted Lie algebras under consideration are derived. The automorphism groups turn out to be the well-known classes of simple linear groups. As an application we obtain by a uniform method a generalization of the known isomorphism theorems between certain of these groups.

From the point of view of abstract Lie algebra theory the present paper contributes only examples of simple restricted Lie algebras of characteristic p. We can not prove, as in the characteristic 0 case, that these are "almost all" (all but the algebras of a finite number of exceptional orders) of the simple restricted Lie algebras with finite bases. In a continuation of this paper we shall consider another infinite class of algebras of this type. It is not known whether or not these together constitute "almost all" of the set of simple restricted Lie algebras with finite bases.

The main part of the discussion is restricted to normal algebras. In the last sections we indicate how the results may be extended to non-normal algebras. The derivation algebras of certain of these simple Lie algebras have the property that the difference algebras $\mathfrak{D}/\mathfrak{I}$, $\mathfrak{I}$ the algebra of inner derivation, are simple. This disproves a conjecture recently made by Zassenhaus.[2] It should be noted, however, that the underlying field for these examples is imperfect so that the question of the validity of Zassenhaus' conjecture for algebras with a finite basis over a perfect field remains open.

---

* Received January 25, 1941.

[1] "Simple Lie algebras over a field of characteristic zero," *Duke Mathematical Journal*, vol. 3 (1938), pp. 534-551. See also the references given there.

[2] "Über Liesche Ringe mit Primzahlcharakteristik," *Abhandl. Math. Sem. Hansischen Univ.*, vol. 13 (1939), p. 80.

0.2. A vector space $\mathfrak{L}$ over a field Φ of characteristic p ($\neq 0$) is a *restricted Lie algebra* if there are defined in addition to the vector operations two functions $[a, b]$ and $a^{[p]}$ in $\mathfrak{L}$ such that the first is bilinear and satisfies

$$[a, b] = -[b, a] \qquad [a, [b, c]] + [b, [c, a]] + [c, [a, b]] = 0$$

and the second satisfies

$$(a\alpha)^{[p]} = a^{[p]}\alpha^p, \qquad [a, b^{[p]}] = [\cdots[a, \overbrace{b]\cdots b]}^{p}$$
$$(a + b)^{[p]} = a^{[p]} + b^{[p]} + s(a, b)$$

where $s(a, b) = s_1(a, b) + \cdots + s_{p-1}(a, b)$ and $(p - i)s_i(a, b)$ is the coefficient of λ^{p-i-1} in

$$[\cdots[a, \overbrace{\lambda a + b], \lambda a + b], \cdots, \lambda a}^{p-1} + b].$$

Subalgebras, ideals, difference algebras, homomorphisms, isomorphisms and automorphisms are defined in the obvious way.[3]

If $\mathfrak{A}$ is an associative algebra over Φ we obtain a restricted Lie algebra $\mathfrak{A}_l$ by defining $[a, b]$ as $ab - ba$ and $a^{[p]}$ as a^p. For example if Φ_n denotes the algebra of $n \times n$ matrices over Φ, Φ_{nl} is the Lie algebra determined by Φ_n. A homomorphism between a restricted Lie algebra $\mathfrak{L}$ and a subalgebra of Φ_{nl} is a representation of $\mathfrak{L}$. Of particular importance is the adjoint representation which associates to a the linear transformation or matrix $x \rightarrow [x, a]$, x variable in $\mathfrak{L}$.

If $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic it is evident that so also are $\mathfrak{A}_l$ and $\mathfrak{B}_l$ and if $\mathfrak{A}$ and $\mathfrak{B}$ are anti-isomorphic with $a \rightarrow b$ an anti-isomorphism between them then $a \rightarrow -b$ is an isomorphism between $\mathfrak{A}_l$ and $\mathfrak{B}_l$. For if $a_1 \rightarrow -b_1$, $a_2 \rightarrow -b_2$, $[a_1, a_2] = a_1a_2 - a_2a_1 \rightarrow -(b_2b_1 - b_1b_2) = [-b_1, -b_2]$ and $a_1^p \rightarrow -(b_1^p) = (-b_1)^p$. If $\mathfrak{A}$ has an anti-automorphism J the set $\mathfrak{S}(\mathfrak{A}, J)$ of skew elements $a^J = -a$ forms a subalgebra of the restricted Lie algebra $\mathfrak{A}_l$. A second anti-automorphism K is *cogredient* to J if $K = S^{-1}JS$ where S is an automorphism in $\mathfrak{A}$. When this holds $a \rightarrow a^S$ induces an isomorphism between $\mathfrak{S}(\mathfrak{A}, J)$ and $\mathfrak{S}(\mathfrak{A}, K)$. In particular if $\mathfrak{A} = \mathfrak{F}_r$ the algebra of $r \times r$ matrices with elements in an involutorial division algebra[4] $\mathfrak{F}$ and if $f \rightarrow \bar{f}$ is an involution in $\mathfrak{F}$ then J defined by $(f_{ij}) \rightarrow u^{-1}(\bar{f}_{ij})'u$ and K such that $(f_{ij})^K = v^{-1}(\bar{f}_{ij})'v$ are cogredient if $v = (\bar{t}'ut)\rho$, where a' denotes the transposed

[3] See the author's paper "Restricted Lie algebras of characteristic p," *Transactions of the American Mathematical Society*, vol. 50 (1941), pp. 15-25.

[4] We are using Albert's terminology: an involution is an anti-automorphism of period 2 and an algebra is involutorial if it has an involution; A. A. Albert, *Structure of Algebras*, New York (1939), p. 151.

of a and ρ is in the center of $\mathfrak{A}$. For, we define $a^S = t^{-1}at$ and may verify that $K = S^{-1}JS$. If $J = S^{-1}JS$, $a \rightarrow a^S$ induces an automorphism in $\mathfrak{S}(\mathfrak{A}, J)$. Hence if g is a *J-orthogonal element* in the sense that $g^J g = gg^J = 1\gamma \neq 0$, γ in Φ then $a \rightarrow g^{-1}ag$ is an automorphism in $\mathfrak{S}(\mathfrak{A}, J)$. The set of J-orthogonal elements forms a group $O(\mathfrak{A}, J)$ under multiplication. We shall call the factor group $O(\mathfrak{A}, J)/\Phi^*$, Φ^* the set of elements $\neq 0$ in Φ, the projective orthogonal group $PO(\mathfrak{A}, J)$.

A *derivation* D in an arbitrary algebra $\mathfrak{A}$ (not necessarily associative) is a linear transformation in $\mathfrak{A}$ such that $(ab)D = (aD)b + a(bD)$. Their totality is a restricted Lie algebra $\mathfrak{D}$. If $\mathfrak{A}$ is itself a restricted Lie algebra we define a *restricted derivation* to be one which satisfies the condition $a^{[p]}D = [\ldots[aD, \overbrace{a]a]\cdots a}^{p-1}]$. These form a sub-algebra $\mathfrak{D}_0$ of $\mathfrak{D}$ which contains as ideal the set of inner derivations $a \rightarrow [a, b]$.[5]

If $\mathfrak{L}$ is a restricted Lie algebra with a finite basis $x_1, \cdots, x_n$ over Φ such that $[x_i, x_j] = \Sigma x_k\gamma_{kij}$, $x_i^{[p]} = \Sigma x_k\mu_{ki}$ the γ's and μ's are constants of multiplication. For any extension P of Φ we may define an algebra $\mathfrak{L}_\mathrm{P}$ using the same constants as for $\mathfrak{L}$ but embracing all $\Sigma x_i\rho_i$, ρ in P. We recall that if $\mathfrak{D}(\mathfrak{D}_0)$ is the derivation (restricted derivation) algebra of $\mathfrak{L}$, $\mathfrak{D}_\mathrm{P}(\mathfrak{D}_{0\mathrm{P}})$ is the corresponding algebra for $\mathfrak{L}_\mathrm{P}$.

$\mathfrak{L}$ is *simple* if it has no proper ideals. If $\mathfrak{L}_\mathrm{P}$ is simple it is evident that $\mathfrak{L}$ is simple. The elements a, b *commute* if $[a, b] = -[b, a] = 0$. The *center* of $\mathfrak{L}$ is the set of elements c such that $[a, c] = 0$ for all a in $\mathfrak{L}$.

I. ALGEBRAS OF TYPE A.

1.1. A lemma on representations. Let $\mathfrak{L}$ be a restricted Lie algebra with a finite basis over Φ and $a \rightarrow A$ a representation of $\mathfrak{L}$ in Φ_{Nl}. Suppose that $\mathfrak{L}$ contains a commutative subalgebra $\mathfrak{H}$ with a basis h_i, $i = 1, \cdots, m$ such that $h_i^p = h_i$ and that $\mathfrak{H}$ is maximal in the sense that the only elements a such that $[h, a] = 0$ for all h in $\mathfrak{H}$ are the elements of $\mathfrak{H}$. If $h_i \rightarrow H_i$ we have $H_i^p = H_i$ and hence we may assume that all of the H_i have diagonal form $\{m_{1i}, m_{2i}, \cdots, m_{Ni}\}$ where $m_{ji} = 0, 1, \cdots, p-1$. Then the general element $h = \Sigma h_i\lambda_i$ of $\mathfrak{H}$ is represented by $H = \{\Lambda_1, \Lambda_2, \cdots, \Lambda_N\}$ where the Λ's are linear forms in the λ's with coefficients $0, 1, \cdots, p-1$ and are called the *weights* of h in the representation. If we apply this to the adjoint representation we obtain a basis $h_1, h_2, \cdots, h_m, e_\alpha, e_\beta \cdots$ where $[e_\alpha, h] = e_\alpha\alpha$, α a linear form $\neq 0$ in the λ's. The α's are the *roots* of h.

[5] See [3] for the proof of this and the results of the next paragraph.

Now suppose that e_α and $e_{-\alpha}$ are contained in the normalized basis. $[e_{-\alpha}, e_\alpha]$ commutes with h since

$$[[e_{-\alpha}, e_\alpha], h] = [[e_{-\alpha}, h], e_\alpha] + [e_{-\alpha}, [e_\alpha, h]] = -[e_{-\alpha}, e_\alpha]\alpha + [e_{-\alpha}, e_\alpha]\alpha = 0.$$

Hence $[e_{-\alpha}, e_\alpha] = h_\alpha \in \mathfrak{H}$. We consider again an arbitrary representation where $h \to H$, $e_\alpha \to E_\alpha$ etc. Suppose that $y_0 \neq 0$ is a vector in the representation space such that $y_0 H = y_0 \Lambda$, Λ one of the Λ_i and $y_0 E_\alpha = 0$. The vector $y_i = y_0 E_{-\alpha}{}^i$ satisfies $y_i H = y_i(\Lambda - i\alpha)$. Hence either all of the forms $\Lambda - i\alpha$, $i = 0, 1, \cdots, p-1$ are weights or there is a k, $0 \leqq k < p-1$ such that $y_k \neq 0$, $y_{k+1} = 0$. By induction we may prove that

$$y_i E_\alpha = y_{i-1}\mu_i; \qquad \mu_i = i\Lambda_\alpha - \frac{i(i-1)}{2}\alpha_\alpha$$

where $\Lambda_\alpha = \Lambda(h_\alpha)$ the value of Λ for h_α and $\alpha_\alpha = \alpha(h_\alpha)$. Since $y_{k+1} = 0$, $y_k \neq 0$ this implies $\mu_{k+1} = 0$.

LEMMA 1. *Let $\mathfrak{L}$ be a restricted Lie algebra with a finite basis and $\mathfrak{H}$ a maximal commutative subalgebra with a basis h_i such that $h_i{}^p = h_i$. Suppose that e_α, $e_{-\alpha}$ are elements of $\mathfrak{L}$ such that $[e_\alpha, h] = e_\alpha \alpha$ for $h = \Sigma h_i \lambda_i$ where α is a linear form $\neq 0$ in the λ's and that $y_0 \neq 0$ is a vector of a representation space such that $y_0 H = y_0 \Lambda$, $y_0 E_\alpha = 0$. Then either all the forms $\Lambda - i\alpha$ are weights of h in this representation or there is a k, $0 \leqq k < p-1$ such that $(k+1)\Lambda_\alpha - \frac{k(k+1)}{2}\alpha_\alpha = 0$, $\Lambda_\alpha = \Lambda(h_\alpha)$, $h_\alpha = [e_{-\alpha}, e_\alpha]$ and $\Lambda - k\alpha$ is a weight.*

1.2. Ideals in Φ_{nl}. We begin with the restricted Lie algebra Φ_{nl}, Φ an infinite field of characteristic p. As usual let e_{ij} denote a matrix basis for Φ_n, i.e. $e_{ij}e_{kl} = \delta_{jk}e_{il}$. Set $h_i = e_{ii}$, $h = \Sigma h_i \lambda_i$ and if $p \neq 2$, $e_{\lambda_i - \lambda_j} = e_{ji}$, $i \neq j$. For $p = 2$ we denote instead e_{ij}, $i < j$, by $e^{(1)}_{\lambda_i + \lambda_j}$ and e_{ji} as $e^{(2)}_{\lambda_i + \lambda_j}$. For $p \neq 2$ we obtain the following multiplication table

$$
\begin{aligned}
&[h_i, h_j] = 0\\
&[e_\alpha, h] = e_\alpha \alpha, \qquad \alpha = \lambda_i - \lambda_j\\
(1)\qquad &[e_\alpha, e_\beta] = \begin{cases} 0 \text{ if } \alpha + \beta \text{ is not a root, } \alpha \neq -\beta\\ \pm e_{\alpha+\beta} \text{ if } \alpha + \beta \text{ is a root, } \alpha \neq -\beta \end{cases}\\
&[e_{-\alpha}, e_\alpha] = h_\alpha = h_i - h_j
\end{aligned}
$$

and

$$(2) \qquad e_\alpha{}^p = 0, \quad h_i{}^p = h_i.$$

We note that $\alpha_\alpha = \alpha(h_\alpha) = 2$. If $p = 2$ we have to replace e_α and $e_{-\alpha}$ by $e_\alpha{}^{(1)}$ and $e_\alpha{}^{(2)}$ and obtain a similar table.

We have shown elsewhere [6] that the only ideals in Φ_{nl} are (1) (the multiples of $1 = \Sigma h_i$) and Φ'_{nl}, the algebra generated by the elements $[a, b]$. The former is the center of Φ_{nl} and the latter may be characterized as the set of elements of trace 0. If $p \nmid n$, $\Phi'_{nl} \wedge (1) = 0$ and $\Phi_{nl} = \Phi'_{nl} \oplus (1)$.

1.3. Automorphisms of Φ_{nl}. We note first that the following mappings are automorphisms: $a \rightarrow a + i \operatorname{tr} a$, $i = 0, 1, 2, \cdots, p-1$, $ni + 1 \not\equiv 0 \pmod p$, $a \rightarrow t^{-1}at$, $a \rightarrow -a'$. The last two are clear. The first follows since $\operatorname{tr} a^p = (\operatorname{tr} a)^p$ and $ni + 1 \not\equiv 0$ insures that $a + i \operatorname{tr} a \neq 0$ if $a \neq 0$. We shall show that these generate the group of automorphisms of Φ_{nl} over Φ. Suppose that $h \rightarrow h^S$, $e_a \rightarrow e_a{}^S$ is an arbitrary automorphism. Then this correspondence is a second representation of Φ_{nl} in Φ_{nl}. Let $\Lambda_1, \Lambda_2, \cdots, \Lambda_n$ be the weights of h in the second representation. If T is the automorphism $a \rightarrow t^{-1}at$ which transforms h^S into the diagonal form $\{\Lambda_1, \Lambda_2, \cdots, \Lambda_n\}$ we have $h^{ST} = \{\Lambda_1, \Lambda_2, \cdots, \Lambda_n\}$.

Since $h_1{}^{ST}, \cdots, h_n{}^{ST}$ are linearly independent the linear forms $\Lambda_1, \Lambda_2, \cdots, \Lambda_n$ are linearly independent mod p. Hence if $\Lambda = \Sigma m_i \lambda_i$ is a weight at most one of the forms $\Lambda + \alpha$, $\Lambda - \alpha$, $\alpha = \lambda_i - \lambda_j$, is a weight and we may suppose first that $\Lambda + \alpha$ is not. Thus if $y_0 \neq 0$, $y_0 H = y_0 \Lambda$ then $y_0 E_a = 0$ and by the above lemma if $p \neq 2$, $\Lambda_a = m_i - m_j = 0, 1$ according as $y_0 E_{-a} = 0$ or $\neq 0$. If $\Lambda - \alpha$ is not a weight but $\Lambda + \alpha$ is, $\Lambda_{-a} = m_j - m_i = 0, 1$. Since $\Lambda - (m_i - m_j)\alpha$ is a weight the weights are invariant under permutation of the λ's. If $p = 2$ and Λ is a weight we have that either $\Lambda + \alpha$ is a weight or for any y_0 such that $y_0 H = y_0 \Lambda$, $y_0 E_a{}^{(i)} = 0$. Then $y_0 H_a = 0$ and $\Lambda_a = m_i + m_j = 0$. Thus in this case also the weights are invariant under permutation of the λ's.

If $\Lambda = m_1(\lambda_1 + \cdots + \lambda_{n_1}) + m_2(\lambda_{n_1+1} + \cdots + \lambda_{n_1+n_2}) + \cdots + m_k(\cdots + \lambda_{n_1 \ldots + n_k})$ where $m_i \neq m_j$ and $k \leq p$ is a weight, we obtain by permuting the λ's $\dfrac{n!}{n_1! \cdots n_k!}$ $(\Sigma n_i = n)$ distinct weights. This number is $> n$ unless $k = 1$ or $k = 2$ and $n_1 = n - 1$, $n_2 = 1$ or $n_1 = 1$, $n_2 = n - 1$. Since there is a weight not of the form $m(\lambda_1 + \cdots + \lambda_n)$ we may suppose $k = 2$, $\Lambda = m_1(\lambda_1 + \cdots + \lambda_{n-1}) + m_2 \lambda_n$ and $m_1 - m_2 = \pm 1$. It follows that $h_1{}^{ST} + \cdots + h_n{}^{ST}$ is the scalar matrix $q = (n-1)m_1 + m_2 = nm_1 \pm 1 \neq 0$.

Consider the automorphism $a \rightarrow a^U = a + i \operatorname{tr} a$. The element h has weights $\Lambda + iq(\lambda_1 + \lambda_2 + \cdots + \lambda_n)$ in the representation $a \rightarrow a^{STU}$. Since $q \neq 0$ we may choose i so that $iq = -m_1$. Then $ni + 1 \not\equiv 0$ and h^{STU} will

---

[6] "Abstract derivation and Lie algebras," *Transactions of the American Mathematical Society*, vol. 42 (1937), pp. 216-217.

be the diagonal matrix $\{-\lambda_1, \cdots, -\lambda_n\}$ or $\{\lambda_1, \cdots, \lambda_n\}$. In the former case we combine with the automorphism $a \to -a' = a^V$ and obtain $h^{STUV} = h$.

Thus in any case we may suppose now that $\bar{S} = STU$ or $= STUV$ has the property $h^{\bar{S}} = h$. Since $[e_\alpha, h] = e_\alpha \alpha$, $[e_\alpha{}^{\bar{S}}, h] = e_\alpha{}^{\bar{S}}\alpha$. If $p \neq 2$ we may suppose that the λ's are chosen so that the roots α are distinct and then $e_\alpha{}^{\bar{S}} = e_\alpha \xi_\alpha$. From $[e_{-\alpha}, e_\alpha] = h_\alpha$, $[e_{-\alpha}{}^{\bar{S}}, e_\alpha{}^{\bar{S}}] = [e_{-\alpha}, e_\alpha]$ and $\xi_{-\alpha}\xi_\alpha = 1$. From $[e_\alpha, e_\beta] = \pm e_{\alpha+\beta}$, $\xi_{\alpha+\beta} = \xi_\alpha + \xi_\beta$. Hence if $\xi_{\lambda_1-\lambda_2} = \cdots = \xi_{\lambda_1-\lambda_n} = 1$, $\xi_{\lambda_2-\lambda_1} = \cdots = \xi_{\lambda_n-\lambda_1} = 1$ and $\xi_\alpha = 1$ for all α. $\bar{S}$ is then the identity automorphism. The automorphism $a \to \bar{h}^{-1} a \bar{h} = a^W$ where $\bar{h} = \Sigma\, h_i \bar{\lambda}_i$ sends h into itself and $e_{\lambda_i-\lambda_j}$ into $e_{\lambda_i-\lambda_j}\bar{\lambda}_j{}^{-1}\bar{\lambda}_i$. We may suppose that $\bar{\lambda}_j{}^{-1}\bar{\lambda}_1 = \xi^{-1}_{\lambda_1-\lambda_j}$. Then $\bar{S}W$ is the identity and hence $\bar{S} = W^{-1}$ has the form $a \to \bar{h} a \bar{h}^{-1}$.

If $p = 2$ we have for $\alpha = \lambda_1 - \lambda_i$

$$(3) \qquad e_\alpha{}^{(1)\bar{S}} = e_\alpha{}^{(1)}\lambda_{11}{}^{(i)} + e_\alpha{}^{(2)}\lambda_{12}{}^{(i)}; \qquad e_\alpha{}^{(2)\bar{S}} = e_\alpha{}^{(1)}\lambda_{21}{}^{(i)} + e_\alpha{}^{(2)}\lambda_{22}{}^{(i)}.$$

Since $(e_\alpha{}^{(1)\bar{S}})^2 = (e_\alpha{}^{(2)\bar{S}})^2 = 0$,

$$(4) \qquad \lambda_{11}{}^{(i)}\lambda_{12}{}^{(i)} = \lambda_{21}{}^{(i)}\lambda_{22}{}^{(i)} = 0.$$

Since $[e_\alpha{}^{(1)}, e_\alpha{}^{(2)}]^{\bar{S}} = [e_\alpha{}^{(1)}, e_\alpha{}^{(2)}]$,

$$(5) \qquad \lambda_{11}{}^{(i)}\lambda_{22}{}^{(i)} + \lambda_{12}{}^{(i)}\lambda_{21}{}^{(i)} = 1.$$

Hence either $\lambda_{11}{}^{(i)} = \lambda_{22}{}^{(i)} = 0$ and $\lambda_{12}{}^{(i)}\lambda_{21}{}^{(i)} = 1$ or $\lambda_{12}{}^{(i)} = \lambda_{21}{}^{(i)} = 0$ and $\lambda_{11}{}^{(i)}\lambda_{22}{}^{(i)} = 1$. From (1)

$$e_{ij}{}^{\bar{S}} = [e^{(2)\bar{S}}_{\lambda_1-\lambda_i}, e^{(1)\bar{S}}_{\lambda_1-\lambda_j}] = e_{ij}\lambda_{22}{}^{(i)}\lambda_{11}{}^{(j)} + e_{ji}\lambda_{21}{}^{(i)}\lambda_{12}{}^{(j)}$$

if $1, i, j$ are unequal and since $[e_{ij}, e_{ji}]^{\bar{S}} = [e_{ij}, e_{ji}]$,

$$(6) \qquad \lambda_{22}{}^{(i)}\lambda_{22}{}^{(j)}\lambda_{11}{}^{(i)}\lambda_{11}{}^{(j)} + \lambda_{21}{}^{(i)}\lambda_{21}{}^{(j)}\lambda_{12}{}^{(i)}\lambda_{12}{}^{(j)} = 1$$

and, hence, if $\lambda_{12}{}^{(i)} = \lambda_{21}{}^{(i)} = 0$, $\lambda_{12}{}^{(j)} = \lambda_{21}{}^{(j)} = 0$. If this holds we may show, as in the case $p \neq 2$, that $\bar{S}$ has the form $a \to \bar{h} a \bar{h}^{-1}$. Otherwise $\lambda_{11}{}^{(i)} = \lambda_{22}{}^{(i)} = 1$, $i = 2, \cdots, n$, and it follows readily that $a^{\bar{S}} = \bar{h} a' \bar{h}^{-1}$. We have therefore proved

LEMMA 2. *Any automorphism of Φ_{nl} is a product of automorphisms of the following types:*

I.	$a \to a + i \operatorname{tr} a$	$(i = 0, 1, \cdots, p-1, ni + 1 \neq 0)$.
II.	$a \to t^{-1}at$	
III.	$a \to -a'$.	

It is interesting to note that if $n \equiv 0 \pmod p$, $\operatorname{tr} 1 = 0$ and hence for any automorphism S, $1^S = \pm 1$.

The automorphisms of the form $a \to a + i \operatorname{tr} a$ form a group. The product of $a \to a + i \operatorname{tr} a$, $a \to a + j \operatorname{tr} a$ is $a \to a + (i + j + nij) \operatorname{tr} a$. Hence if $n \equiv 0 \pmod p$ this group is isomorphic to the additive group mod p and if $n \not\equiv 0 \pmod p$ the correspondence between $a \to a + i \operatorname{tr} a$ and $ni + 1$ is an isomorphism with the multiplicative group mod p. The automorphisms I and II form invariant subgroups. Any element in I commutes with any in II. If $n = 2$, $a' = -q^{-1}aq + \operatorname{tr} a$ if $q = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ so that I and II generate the whole group of automorphisms. If $n > 2$ the automorphisms I and II form an invariant subgroup of index 2. For otherwise we should have a matrix s such that $a' = -s^{-1}as - i \operatorname{tr} a$. For a of trace 0 we obtain $a = s's^{-1}as(s')^{-1}$. Hence $s's^{-1}$ commutes with all matrices of trace 0 and therefore with all matrices in Φ_n. It follows that $s's^{-1} = 1\rho$, $s = s\rho$ and since $s'' = s$, $s' = \pm s$. As we shall show in part II, if s is symmetric or skew the dimensionality of the space of matrices a such that $s^{-1}a's = -a$ is $n(n-1)/2$ or $n(n+1)/2$ (n even) and these numbers are $< n^2 - 1$ the dimensionality of the space of matrices of trace 0 if $n > 2$. Hence the equation $a' = -s^{-1}as - i \operatorname{tr} a$ can not hold for a fixed s and all a.

1.4. Restricted derivations of Φ_{nl}. Let D be a restricted derivation. We choose the $\lambda_i = \bar{\lambda}_i$ in $h = \Sigma h_i\lambda_i$ such that all of the p^n linear forms $\Sigma m_i\bar{\lambda}_i$, $m_i = 0, \cdots, p-1$ are distinct. Then the minimum p-polynomial satisfied by $\bar{h} = \Sigma h_i\bar{\lambda}_i$ is $\Pi(\lambda - \Sigma m_i\bar{\lambda}_i) = \lambda^{p^n} + \lambda^{p^{n-1}}\beta_{n-1} + \cdots + \lambda\beta_0$ where $\beta_0 = \prod_{\neq 0} (\Sigma m_i\bar{\lambda}_i) \neq 0$.[7] Thus

$$\bar{h}^{p^n} + \bar{h}^{p^{n-1}}\beta_{n-1} + \cdots + \bar{h}\beta_0 = 0.$$

Since D is restricted

$$(\bar{h}D)(\bar{H}^{p^n-1} + \bar{H}^{p^{n-1}-1}\beta_{n-1} + \cdots + \beta_0) = 0$$

where $\bar{H}$ denotes the operation $a \to [a, \bar{h}]$.

$$\bar{H}^{p^n-1} + \cdots + \beta_0 = \prod_{\neq 0} (\bar{H} - \Sigma m_i\bar{\lambda}_i).$$

Hence if we set $\bar{h}D = h_0 + \Sigma e_\alpha\sigma_\alpha$ we obtain $h_0 = 0$, $\bar{h}D = \Sigma e_\alpha\sigma_\alpha$. Set $\tau_\alpha = -\bar{\alpha}^{-1}\sigma_\alpha$ and subtract from D the derivation $a \to [a, \Sigma e_\alpha\tau_\alpha]$. The resulting derivation E has the property $\bar{h}E = 0$. Since $\bar{h}, \bar{h}^p, \cdots, \bar{h}^{p^{n-1}}$ are

---

[7] Cf. Ore "On a special class of polynomials," *Transactions of the American Mathematical Society*, vol. 35 (1933), pp. 564-565.

linearly independent they form a basis for the subalgebra $\mathfrak{H} = (h_1, h_2, \cdots, h_n)$. Since $(\bar{h}^{p^i})E = 0$, $h_i E = 0$. Now $[e_a, \bar{h}]E = [e_a E, \bar{h}] = (e_a E)\bar{\alpha}$ implies that for $p \neq 2$, $e_a E = e_a \xi_a$ and for $p = 2$, $e_a{}^{(1)}E = e_a{}^{(1)}\lambda_{11} + e_a{}^{(2)}\lambda_{12}$, $e_a{}^{(2)}E = e_a{}^{(1)}\lambda_{21} + e_a{}^{(2)}\lambda_{22}$. In the former case it follows readily that E, and hence D, is inner. In case $p = 2$ we use $(e_a{}^{(k)})^2 = 0$, $[e_a{}^{(k)}E, e_a{}^{(k)}] = 0$ and hence $e_a{}^{(k)}E = e_a{}^{(k)}\xi_a{}^{(k)}$ and we can prove that E is inner as in the case $p \neq 2$.

LEMMA 3. *Any restricted derivation of Φ_{nl} is inner.*

1.5. Restricted Lie algebras of type A. Let $\mathfrak{A}$ be a normal simple associative algebra of order n^2 over an arbitrary Φ of characteristic p and $\mathfrak{A}_l$ the restricted Lie algebra determined by $\mathfrak{A}$. $\mathfrak{A}_l$ will be called a restricted Lie algebra of *type* A_1. If Ω is the algebraic closure of Φ, $\mathfrak{A}_\Omega = \Omega_n$ and hence $\mathfrak{A}_{l\Omega} = \Omega_{nl}$. The algebra $\mathfrak{A}'_l$ generated by elements of the form $[a, b]$ is an ideal of order $n^2 - 1$ in $\mathfrak{A}_l$ since $(\mathfrak{A}'_l)_\Omega = \Omega'_{nl}$ is an ideal of this order in Ω_{nl}. We may choose a basis $a_1, \cdots, a_{n^2}$ for $\mathfrak{A}$ over Φ such that $a_1, \cdots, a_{n^2-1}$ is a basis for $\mathfrak{A}'_l$. Then these elements will constitute bases for Ω_n and Ω'_{nl} also, i. e. every element of Ω_n has the form $\sum_{j=1}^{n^2} a_i\omega_i$, $\omega \epsilon \Omega$ and every element of Ω'_{nl} has the form $\sum_{j=1}^{n^2-1} a_j\omega_j$. It is well known that the characteristic polynomial of any element $a = \sum a_i\phi_i$ of $\mathfrak{A}$ has coefficients in Φ. In particular $\operatorname{tr} a \,\epsilon\, \Phi$ and the correspondence $a \to a + i \operatorname{tr} a$ is an automorphsim of $\mathfrak{A}_l$ if $i = 0, 1, \cdots, p - 1$ and $ni + 1 \not\equiv 0 \pmod{p}$.

Now suppose that $\mathfrak{A}$ is a simple associative algebra of order n^2 over its center $\mathrm{P} = \Phi(q)$ a quadratic extension of Φ and suppose that $\mathfrak{A}$ has an involution J of second kind.[8] Let $\mathfrak{S}(\mathfrak{A}, J)$ be the set of J-skew elements. It is known that there are n^2 elements $a_1, a_2, \cdots, a_{n^2}$ such that the elements of $\mathfrak{S}(\mathfrak{A}, J)$ are the sums $\sum a_i\phi_i$, $\phi \,\epsilon\, \Phi$ and those of $\mathfrak{A}$ are $\sum a_i\rho_i$, ρ_i in P.[9] $\mathfrak{S}(\mathfrak{A}, J)$ is a restricted Lie algebra over Φ. Since $\mathfrak{S}(\mathfrak{A}, J)_\mathrm{P} = \mathfrak{A}_l$, $\mathfrak{S}(\mathfrak{A}, J)_\Omega = \Omega_{nl}$ if Ω is the algebraic closure of P. It follows as above that if $\mathfrak{S}(\mathfrak{A}, J)'$ denotes the space generated by the $[a, b]$ a, b in $\mathfrak{S}$, $\mathfrak{S}(\mathfrak{A}, J)'$ is a restricted Lie algebra such that $\mathfrak{S}'_\Omega = \Omega'_{nl}$. Hence we may suppose that $a_1, \cdots, a_{n^2-1}$ form a basis for $\mathfrak{S}'$ over Φ, for $\mathfrak{A}'_l$ over P and for Ω'_{nl} over Ω. Let $\Gamma = \Phi(\xi_1, \cdots, \xi_{n^2})$ be the field obtained from Φ by adjoining the indeterminates ξ_i and consider the algebra $(\mathfrak{A}$ over $\Phi)_\Gamma$. The involution J has a unique extension J to $(\mathfrak{A}$ over $\Phi)_\Gamma$. The element $a_\xi = \sum a_i\xi_i$ is skew relative to this involution.

[8] An involution J is of first kind or second kind according as it induces the identity automorphism or not in the center P. If J is of second kind then P is separable over Φ the field of invariant elements of P. Cf. Albert, *loc. cit.* [4], p. 153.

[9] Albert [4], p. 153.

We have

$$(7) \qquad (a_\xi)^n - (a_\xi)^{n-1} \operatorname{tr} a_\xi + \cdots = 0$$

and if we operate with J we obtain

$$(8) \qquad (-a_\xi)^n - (-a_\xi)^{n-1} (\operatorname{tr} a_\xi)^J + \cdots = 0.$$

Since (7) is the minimum equation of $\mathfrak{A}$ over P the coefficients of (7) and (8) differ by the sign $(-1)^n$.[10] Hence $(\operatorname{tr} a_\xi)^J = -\operatorname{tr} a_\xi$ and if $a = \Sigma a_i \phi_i$ is any element of $\mathfrak{S}(\mathfrak{A}, J)$, $\operatorname{tr} a \in \mathfrak{S}(\mathfrak{A}, J)$. It follows that $a \to a + i \operatorname{tr} a$ is an automorphism of $\mathfrak{S}$ over Φ if $i = 0, \cdots, p-1$ and $ni + 1 \not\equiv 0 \pmod p$. We shall call $\mathfrak{S}$ a restricted Lie algebra of *type* A_{II}.

1. 6. The enveloping algebras. Consider an algebra $\mathfrak{S}$ of type A_{II} with $p \neq 2$. We may suppose that one of the $a_i = q$, $q^J = -q$, $q^2 = \mu$. Then $a_1 q, a_2 q, \cdots, a_{n^2} q$ are symmetric and together with $a_1, \cdots, a_{n^2}$ form a basis for $\mathfrak{A}$ over Φ. Thus the enveloping algebra over Φ of $\mathfrak{S}(\mathfrak{A}, J)$ is $\mathfrak{A}$. We wish to show that if $n > 2$, $\mathfrak{A}$ is the enveloping algebra of $\mathfrak{S}(\mathfrak{A}, J)'$. Suppose first that Φ is finite. There are no non-commutative division algebras over Φ and hence $\mathfrak{A} = \mathrm{P}_n$, $\mathrm{P} = \Phi(q)$. If $\alpha \to \bar{\alpha}$ is the automorphism in P we may suppose that J is the correspondence $a = (\alpha_{ij}) \to d^{-1} \bar{a}' d$, $\bar{a}' = (\bar{\alpha}_{ji})$, $d = \Sigma e_{ii} \delta_i$, δ_i in Φ.[11] Then the condition that $a \in \mathfrak{S}(\mathfrak{A}, J)$ is that $\bar{\alpha}_{ji} = -(\delta_i / \delta_j) \alpha_{ij}$. Hence $e_{ii} q$, $e_{ij} \delta_j - e_{ji} \delta_i$, $(e_{ij} \delta_j + e_{ji} \delta_i) q$, $i < j$, is a basis for $\mathfrak{S}(\mathfrak{A}, J)$. The elements $(e_{ii} - e_{jj}) q$, $e_{ij} \delta_j - e_{ji} \delta_i$, $(e_{ij} \delta_j + e_{ji} \delta_i) q \in \mathfrak{S}(\mathfrak{A}, J)'$. If $n > 2$ choose $k \neq i, j$. Then $(e_{ii} - e_{kk}) q (e_{ij} \delta_j - e_{ji} \delta_i) = e_{ij} q \delta_j$ is in the enveloping algebra $\mathfrak{B}$. Similarly e_{ij}, e_{ji}, $e_{ji} q \in \mathfrak{B}$. Hence also e_{ii}, $e_{ii} q \in \mathfrak{B}$ and $\mathfrak{B} = \mathfrak{A}$. Now suppose that Φ is infinite. It suffices to show that the enveloping algebra $\mathfrak{B} \geq \mathfrak{S}(\mathfrak{A}, J)$. Since if $a \in \mathfrak{S}(\mathfrak{A}, J), a^3, a^5, \cdots \in \mathfrak{S}(\mathfrak{A}, J)$ we must show that there is an a in $\mathfrak{S}(\mathfrak{A}, J)'$ such that a suitable $a^{2m+1} \notin \mathfrak{S}(\mathfrak{A}, J)'$. If this is not the case we should have for all m and ϕ_j in Φ, $\operatorname{tr} \left(\sum_1^{n^2-1} a_j \phi_j\right)^{2m-1} = 0$. Since Φ is infinite this implies that $\operatorname{tr} (\Sigma a_j \omega_j)^{2m-1} = 0$ and this is impossible if $n > 2$ since there are matrices of trace 0, say $\Sigma e_{ii} \omega_i$, $\Sigma \omega_i = 0$, such that $\Sigma \omega_i^{2m-1} \neq 0$. We need merely choose $2m - 1 \neq p^e$ and then satisfy the condition since $-(\omega_2 + \cdots + \omega_n)^{2m-1} + \omega_2^{2m-1} + \cdots + \omega_n^{2m-1} \not\equiv 0$.

Now let $p = 2$. We may suppose that $q^J = q + 1$. If $a_1, \cdots, a_{n^2}$ is a basis for $\mathfrak{S}(\mathfrak{A}, J)$, $a_1 q, \cdots, a_{n^2} q$ is a basis for $\mathfrak{S} q$ and $\mathfrak{A} = \mathfrak{S} + \mathfrak{S} q$

---

[10] Albert [4], p. 123.

[11] The involutions of second kind in P_n are of the form $a \to s^{-1} \bar{a}' s$ where $\bar{s}' = s$. Hence if the s associated with J is not diagonal we replace J by the cogredient involution $a \to d^{-1} \bar{a}' d$ where $d = \bar{g}' s g$ is diagonal.

and $\mathfrak{S} \triangle \mathfrak{S}q = 0$. Hence if the enveloping algebra $\mathfrak{B}$ over Φ of $\mathfrak{S}$ contains an element not in $\mathfrak{S}$ it will contain an element b in $\mathfrak{S}q$. If $\mathfrak{A}$ is a division algebra choose r, s in $\mathfrak{S}$ such that $[r, s] \neq 0$. Then $(rs)^J = sr \neq rs$ and hence $\mathfrak{B}$ contains an element aq, $a \neq 0$, in $\mathfrak{S}$. Then $a^{-1} \epsilon \mathfrak{S}$ and $\mathfrak{B}$ contains q and therefore $\mathfrak{B} = \mathfrak{A}$.

If $\mathfrak{A}$ is not a division algebra $\mathfrak{A} = \mathfrak{F}_r$ where $r > 1$ and $\mathfrak{F}$ is a division algebra (commutative or not). We may suppose that the involution in $a \rightarrow d^{-1}\bar{a}'d$ where $a = (a_{ij})$, $\bar{a}' = (\bar{a}_{ji})$ and $a_{ji} \rightarrow \bar{a}_{ji}$ is an involution of the second kind in $\mathfrak{F}$ and $d = \Sigma e_{ii}d_i$, $\bar{d}_i = d_i$.[12] The condition that $a \epsilon \mathfrak{S}(\mathfrak{A}, J)$ is $d_i a_{ij} = \bar{a}_{ji} d_j$. Hence if $\mathfrak{F} = \mathrm{P}$ the elements $e_{ii}, e_{ij}d_j + e_{ji}d_i$ and $e_{ij}d_jq + e_{ji}d_i(q+1) \epsilon \mathfrak{S}$. Thus $e_{ii}(e_{ij}d_j + e_{ji}d_i) = e_{ij}d_j \epsilon \mathfrak{B}$ and since $d_j \epsilon \Phi$, $e_{ij} \epsilon \mathfrak{B}$. Similarly $e_{ij}q \epsilon \mathfrak{B}$ if $i \neq j$ and hence $e_{ii}, e_{ii}q \epsilon \mathfrak{B}$ so that $\mathfrak{B} = \mathfrak{A}$. Finally suppose $\mathfrak{A} = \mathfrak{F}_r$ where $\mathfrak{F}$ is a non-commutative division algebra. The elements $e_{ii}a_{ii}$ where $d_i^{-1}\bar{a}_{ii}d_i = a_{ii}$ are in $\mathfrak{S}$. Since $a \rightarrow d_i^{-1}\bar{a}d_i$ is an involution of the second kind in $\mathfrak{F}$, $\mathfrak{B}$ includes every $e_{ii}b_{ii}$, b_{ii} arbitrary in $\mathfrak{F}$, and hence every b in $\mathfrak{F}$. Since $\mathfrak{B}$ contains $e_{ij}d_j + e_{ji}d_i$ it follows that $\mathfrak{B}$ contains all $e_{ij}b$ and $\mathfrak{B} = \mathfrak{A}$. As in the case $p \neq 2$ we may prove that $\mathfrak{A}$ is also the enveloping algebra of $\mathfrak{S}(\mathfrak{A}, J)'$ if $n > 2$.

LEMMA 4. *If $\mathfrak{A}$ is a simple associative algebra with center $\mathrm{P} = \Phi(q)$ a quadratic field and $\mathfrak{A}$ has an involution J of second kind, then the enveloping algebra over Φ of $\mathfrak{S}(\mathfrak{A}, J)$ is $\mathfrak{A}$. The same result holds for $\mathfrak{S}(\mathfrak{A}, J)'$ if $n > 2$.*

From now on we suppose that $n > 2$ for all Lie algebras of type A_{II}.

1.7. Isomorphism and automorphisms. Suppose $\mathfrak{L}_1$ and $\mathfrak{L}_2$ are isomorphic restricted Lie algebras of type A. We may suppose that $\mathfrak{L}_1$ is the set $\Sigma a_i\phi_i$, $\mathfrak{L}_2$ the set $\Sigma a_i^S\phi_i$ where S is the isomorphism and Ω_{nl} is the set $\Sigma a_i\omega_i$ or the set $\Sigma a_i^S\omega_i$. Thus the correspondence $\Sigma a_i\omega_i \rightarrow \Sigma a_i^S\omega_i$ is an automorphism in Ω_{nl} which extends the isomorphism between $\mathfrak{L}_1$ and $\mathfrak{L}_2$. Hence there exists an element g in Ω_n such that either $a_i^S = g^{-1}a_ig + j\,\mathrm{tr}\,a_i$, $j = 0, 1, \cdots, p-1$, $jn + 1 \neq 0$ or $a_i^S = -g^{-1}a'_ig + j\,\mathrm{tr}\,a_i$, $jn - 1 \neq 0$. If we replace a_i by $a_i \pm j\,\mathrm{tr}\,a_i = a_i^U$ where U is an automorphism in $\mathfrak{L}_1$ we may suppose $a_i^S = g^{-1}a_ig$ or $a_i^S = -g^{-1}a'_ig$. In the first case the correspondence S determines an isomorphism between the enveloping algebras over

[12] This is a consequence of the theorem that any hermitian matrix in $\mathfrak{F}_r$ is cogredient to a diagonal matrix. The proof is exactly as in [1] pp. 542-544 except for the proof of the Lemma that for any hermitian form $f = (x, y) \neq 0$ there is a vector u such that $(u, u) \neq 0$. This is seen as follows: If $(u, u) = 0$ for all u, $(u + v, u + v) = (u, u) + (v, v) + (u, v) + (v, u) = 0$. Hence $(u, v) = (v, u) = \overline{(u, v)}$ and this implies that $d \rightarrow \bar{d}$ is the identity transformation.

Φ of $\mathfrak{L}_1$ and of $\mathfrak{L}_2$ and in the second case $-S$ is an anti-isomorphism. If $\mathfrak{L}_1$ has type A_I the center of the enveloping algebra is Φ and if it has type A_{II} the center is $\mathrm{P} = \Phi(q)$. Hence we have

THEOREM 1. *A restricted Lie algebra of type A_I is not isomorphic to one of type A_{II}.*

By the same method we may prove as in the case of Φ of characteristic 0 the following theorems.

THEOREM 2. *The restricted Lie algebras $\mathfrak{A}_l$ and $\mathfrak{B}_l$ of type A_I are isomorphic if and only if the associative algebras $\mathfrak{A}$ and $\mathfrak{B}$ are either isomorphic or anti-isomorphic.*

THEOREM 3. *The restricted Lie algebras $\mathfrak{S}(\mathfrak{A}, J)$ and $\mathfrak{S}(\mathfrak{B}, K)$ of type A_{II} are isomorphic if and only if $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic and J and K are cogredient.*

THEOREM 4. *Any automorphism of $\mathfrak{A}_l$ has either the form $a \to t^{-1}at + i \operatorname{tr} a$, $i = 0, \cdots, p-1$, $ni + 1 \neq 0$, or $a \to -t^{-1}a't + i \operatorname{tr} a$, $ni - 1 \neq 0$ where $a \to a'$ is an anti-automorphism of $\mathfrak{A}$ over Φ.*

Of course the latter set need not exist.

THEOREM 5. *Any automorphism of $\mathfrak{S}(\mathfrak{A}, J)$ has the form $a \to a^S + i \operatorname{tr} a$ $i = 0, \cdots, p-1$, $ni + 1 \neq 0$, where S is an automorphism of $\mathfrak{A}$ over Φ which commutes with J.*

There are two types of S's: those which alter elements of P and those which are automorphisms of the normal simple algebra $\mathfrak{A}$ over P. The latter are inner and hence have the form $a \to s^{-1}as$ where $ss^J = 1\sigma$, σ in Φ.

By Lemma 3 and the property $\mathfrak{A}_{l\Omega} = \Omega_{nl}$, $\mathfrak{S}_\Omega = \Omega_{nl}$ we have

THEOREM 6. *Any restricted derivation of a restricted Lie algebra of type A is inner.*

1.8. The derived algebras. If $\mathfrak{L}$ has type A, $\mathfrak{L}_\Omega = \Omega_{nl}$ and $\mathfrak{L}'_\Omega = \Omega'_{nl}$. Hence by a previous result [13] we have

THEOREM 7. *If $\mathfrak{L}$ is a restricted Lie algebra of type A, $\mathfrak{L}'$ is simple if $p \nmid n$ and $\mathfrak{L}'/(1)$, (1) the set 1α, is simple if $p \mid n$ except when $p = n = 2$.*

By simplicity here we mean that there are no (restricted) ideals. As a

---

[13] *Loc. cit.*[6]

matter of fact this theorem holds when $\mathfrak{L}'$, $\mathfrak{L}'/(1)$ are regarded as ordinary Lie algebras, i. e. the algebras have no ordinary ideals other than (0) and the whole algebra.

In the remainder of this section we suppose $p \nmid n$. Then $\Omega_{ln} = \Omega'_{nl} \oplus (1)$. If $a \rightarrow a^S$ is an automorphism in Ω'_{nl} we define $1^S = 1$ and obtain an automorphism in Ω_{nl}. By Lemma 2 we have either $a^S = t^{-1}at$ or $a^S = -t^{-1}a't$ where $t \in \Omega_{nl}$. If we use the second part of Lemma 4 we obtain

THEOREM 1'. *The derived algebra of a restricted Lie algebra of type A_I is not isomorphic to the derived algebra of one of type A_{II} if $n > 2$ and $\not\equiv 0 \pmod p$.*

Similarly Theorems 2 and 3 hold with $\mathfrak{A}_l, \mathfrak{B}_l, \mathfrak{S}(\mathfrak{A}, J), \mathfrak{S}(\mathfrak{A}, K)$ replaced by their derived algebras. Theorems 4 and 5 hold when this substitution is made and $i = 0$. If we note that any restricted derivation in Ω'_{nl} may be extended to Ω_{nl} by defining $1D = 0$ we obtain the fact that the restricted derivations of any $\mathfrak{A}'_l$, $\mathfrak{S}'$ with $p \nmid n$, are all inner. Finally we remark that since the algebras $\mathfrak{A}'_l$, $\mathfrak{S}'$ have no centers all of these results hold when we regard these algebras as ordinary Lie algebras. Thus for example if $a \rightarrow a^S$ is a correspondence in $\mathfrak{A}'_l$ such that $(a+b)^S = a^S + b^S$, $(a\alpha)^S = a^S\alpha$, $[a,b]^S = [a^S, b^S]$ then there is a t in $\mathfrak{A}$ such that either $a^S = t^{-1}at$ or $a^S = -t^{-1}a't$, $a \rightarrow a'$ an anti-automorphism. In the same way we may drop the word restricted in the statement of Theorem 6 for the derived algebra.

1.9. The completeness theorem. We suppose that $\mathfrak{L}'$ is a restricted Lie algebra such that $\mathfrak{L}'_\Omega = \Omega'_{nl}$ and $p \nmid n$. Thus $\mathfrak{L}'$ is a subset of Ω'_{nl} and contains $n^2 - 1$ elements $a_1, a_2, \cdots, a_{n^2-1}$ such that every element of $\mathfrak{L}$ has the form $\Sigma\, a_j\phi_j$, ϕ in Φ and every element of Ω'_{nl} may be written in one and only one way as $\Sigma\, a_j\omega_j$, ω in Ω. The coefficients of the matrices a_j generate a finite algebraic extension Γ of Φ. Thus the a_j are linear combinations of $e_{ii} - e_{jj}$ and e_{ij}, $i \neq j$, with coefficients in Γ and hence the $e_{ii} - e_{ji}$, e_{ij} are linear combinations of the a_j with coefficients in Γ also. It follows that the set $\Sigma\, a_j\gamma_j$, γ in Γ, is Γ'_{nl} and so we may replace the algebraically closed field Ω by the finite extension Γ in our discussion.

LEMMA 5. *If $\mathfrak{L}$ is a restricted Lie algebra of type A over Φ there exists a separable extension Σ of Φ such that $\mathfrak{L}_\Sigma = \Sigma_{nl}$ and $\mathfrak{L}'_\Sigma = \Sigma'_{nl}$.*

From the theory of associative algebras we know that every normal simple algebra $\mathfrak{A}$ has a separable splitting field Σ, i. e., $\mathfrak{A}_\Sigma = \Sigma_n$. Hence if $\mathfrak{L} = \mathfrak{A}_l$, $\mathfrak{L}_\Sigma = \mathfrak{A}_{l\Sigma} = \Sigma_{nl}$. If $\mathfrak{L} = \mathfrak{S}(\mathfrak{A}, J)$, $\mathfrak{A}$ with center $P = \Phi(q)$, $\mathfrak{L}_P = \mathfrak{A}_{lP}$.

There is a field Σ separable over P such that $\mathfrak{L}_\Sigma = \mathfrak{A}_{l\Sigma} = \Sigma_{nl}$. Since Σ is separable over Φ it has the required properties.

We may now prove the following "completeness" theorem.

THEOREM 8. *If $\mathfrak{L}'$ is a restricted Lie algebra over Φ such that $\mathfrak{L}'_\Omega = \Omega'_{nl}$ and $p \nmid n$ then either there exists a normal simple associative algebra $\mathfrak{A}$ such that $\mathfrak{L}' \cong \mathfrak{A}'_l$ or there exists a simple associative algebra with center $P = \Phi(q)$, a quadratic field over Φ, and with an involution J of the second kind such that $\mathfrak{L}' \cong \mathfrak{S}(\mathfrak{A}, J)'$.*

We suppose that Γ is chosen as above and let Δ be the maximal separable subfield of Γ. Then there is a chain of subfields $\Delta_0 = \Delta$, $\Delta_i = \Delta_{i-1}(x_i)$, $x_i^p = \xi_i \in \Delta_{i-1}$ and $\Delta_u = \Gamma$. Suppose first that $u = 0$, i. e. that Γ is separable. Then we may extend Γ to $\bar{\Gamma}$ which is normal and separable over Φ and we have $\mathfrak{L}'_{\bar{\Gamma}} = \bar{\Gamma}'_{nl}$. Let $\mathfrak{G} = (1, s, \cdots, v)$ be its Galois group. If $a = (\alpha_{ij}) \in \bar{\Gamma}'_{nl}$ we define $a^s = (\alpha_{ij}{}^s)$ and if $a = \Sigma\, a_j\bar{\gamma}_j$ define $a^{S_1} = \Sigma\, a_j{}^s\bar{\gamma}_j$. Since $[a_j, a_k] = \Sigma\, a_p\alpha_{pjk}$, α_{pjk} in Φ, $a_j{}^p = \Sigma\, a_q\beta_{qj}$, β_{qj} in Φ, S_1 is an automorphism of $\bar{\Gamma}'_{nl}$ over $\bar{\Gamma}$. Hence there exists a t_s in $\bar{\Gamma}_n$ such that either $a^{S_1} = t_s{}^{-1}at_s$ or $a^{S_1} = -t_s{}^{-1}a't_s$. The s for which S_1 is of the first type form an invariant subgroup $\mathfrak{h}$ of index 1 or 2 in $\mathfrak{g}$ since the automorphisms of the form II form such a subgroup in the group of all automorphisms of $\bar{\Gamma}'_{nl}$ over $\bar{\Gamma}$.[14] Suppose first that $\mathfrak{h} = \mathfrak{g}$. Then the mappings S defined by $a^S = a^{sS_1^{-1}}$ generate automorphisms of the associative algebra $\bar{\Gamma}_n$ over Φ. The elements left invariant by the S form, therefore, an algebra $\mathfrak{A}$ over Φ. Since $a_1, \cdots, a_{n^2-1}$, $a_{n^2} = 1$, is a basis for $\bar{\Gamma}_n$ and $(\Sigma\, a_i\bar{\gamma}_i)^S = \Sigma\, a_i\bar{\gamma}_i{}^s$, $\mathfrak{A}$ consists of the elements $\Sigma\, a_i\phi_i$, ϕ_i in Φ. The elements $a_1, \cdots, a_{n^2-1}$ are linear combinations of commutators $[a_j, a_k]$ and therefore the set $\sum_1^{n^2-1} a_j\phi_j$ is $\mathfrak{A}'_l$. If $\mathfrak{h}$ has index 2 there is a quadratic subfield $P = \Phi(q)$ such that the Galois group of $\bar{\Gamma}$ over P is $\mathfrak{h}$. Then the set $\Sigma\, a_i\rho_i$, ρ_i in P, $a_{n^2} = 1$ is a simple algebra with P as center. Let v be an element in $\mathfrak{g}$ not in $\mathfrak{h}$ and $a_j{}^{V_1} = -t_v{}^{-1}a'_jt_v$. v induces the fundamental automorphism $\rho \to \bar{\rho}$ in P. Consider the correspondence J defined as $a^J = (\Sigma\, a_i\rho_i)^J = -\Sigma\, a_j\bar{\rho}_j + a_{n^2}\bar{\rho}_{n^2}$. Since $a^J = (t_v{}^{v^{-1}})^{-1}(a')^{v^{-1}}t_v{}^{v^{-1}}$, J is an anti-automorphism in $\mathfrak{A}$. By its form it is involutorial and its skew elements are $\Sigma\, a_j\phi_j + (a_{n^2}q)\phi$, $\bar{q} = -q$ in P.[15] Hence $\mathfrak{S}(\mathfrak{A}, J)'$ is the set $\Sigma\, a_j\phi_j = \mathfrak{L}'$.

Now suppose that the theorem holds for $u - 1$ and consider $\mathfrak{L}'_{\Delta_{u-1}}$. This

---

[14] Cf. the author's paper "Simple Lie algebras of type A," *Annals of Mathematics*, vol. 39 (1938), p. 186.

[15] If $p = 2$ we may take $q = 1$.

algebra becomes Γ'_{nl} when Δ_{u-1} is extended to $\Gamma = \Delta_{u-1}(x)$, $x = x_u$, $x^p = \xi \in \Delta_{u-1}$. Let d be a derivation in Γ over Δ_{u-1} such that the elements of Δ_{u-1} are the only d-constants.[16] Then $a = (\alpha_{ij}) \to a^d = (\alpha_{ij}{}^d)$ and $(\Sigma\, a_j\gamma_j)D_1 = \Sigma\, a_j{}^d\gamma_j$ are derivations in Γ_n over Δ_{u-1} and in Γ_{nl} over Γ, respectively. The last statement holds since the multiplication table $[a_j, a_k]$, $a_j{}^p$ has coefficients in Δ_{u-1}. D_1 is inner and, hence, is induced by a derivation of the associative algebra Γ_n over Γ. Hence $D = D_1 - d$ is a derivation in Γ_n over Δ_{u-1} and the set of D-constants is an algebra $\mathfrak{B}$ over Δ_{u-1}. Since $(\Sigma\, a_i\gamma_i)(D_1 - d) = (\Sigma\, a_i\gamma_i)D_1 - (\Sigma\, a_i{}^d\gamma_i + \Sigma\, a_i\gamma_i{}^d) = -\Sigma\, a_i\gamma_i{}^d$ the D-constants are the elements of the form $\Sigma\, a_i\delta_i$, δ_i in Δ_{u-1}. Hence $\mathfrak{L}'_{\Delta_{u-1}} = \mathfrak{B}'_l$. There is a separable extension Λ of Δ_{u-1} such that $\mathfrak{L}'_\Lambda = \mathfrak{B}'_{l\Lambda} = \Lambda'_{nl}$. As is well known $\Lambda = \bar{\Delta}_{u-1}$ where $\bar{\Delta}$ is the maximal separable subfield of Λ over Φ.[17] Hence by the hypothesis of the induction $\mathfrak{L}'$ has the form $\mathfrak{A}'_l$ or the form $\mathfrak{S}(\mathfrak{A}, J)'$.

II. RESTRICTED LIE ALGEBRAS OF TYPES *B*, *C*, *D*.

2.1. Involutions in Ω_n. Since the automorphisms of the associative algebra Ω_n are all inner, the involutions in Ω_n have the form $a^J = s^{-1}a's$ where $s' = \pm s$. We assume that Ω is algebraically closed. Then by replacing J, if necessary, by a cogredient involution we may suppose when $p \neq 2$ that s has one of the following forms

$$(9) \qquad s_1 = \begin{pmatrix} 0 & 1_\nu \\ -1_\nu & 0 \end{pmatrix}; \qquad s_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1_\nu \\ 0 & 1_\nu & 0 \end{pmatrix}; \qquad s_3 = \begin{pmatrix} 0 & 1_\nu \\ 1_\nu & 0 \end{pmatrix}$$

where 1_ν is the identity matrix of ν rows and columns. The corresponding involutions will be denoted as J_1, J_2, J_3. A computation shows that $\mathfrak{S}(\Omega_n, J_i)$, $i = 1, 2, 3$, consists respectively of the following sets of matrices

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad a_{ij} \in \Omega_\nu, \quad a'_{11} = -a_{22}, \quad a'_{12} = a_{12}, \quad a'_{21} = a_{21}$$

$$\begin{pmatrix} 0 & v_1 & v_2 \\ -v'_2 & a_{11} & a_{12} \\ -v'_1 & a_{21} & a_{22} \end{pmatrix}, \quad a_{ij} \in \Omega_\nu, \quad a'_{11} = -a_{22}, \quad a'_{12} = -a_{12}, \quad a'_{21} = -a_{21}$$

and where $v_1 = (b_1, b_2, \cdots, b_\nu)$ and $v_2 = (b_{\nu+1}, \cdots, b_{2\nu})$

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad a'_{11} = -a_{22}, \quad a'_{12} = -a_{12}, \quad a'_{21} = -a_{21}.$$

---

[16] The existence of such derivations has been proved by Baer, "Algebraische Theorie der differentiierbaren Funktionenkörper. I," *Sitzungsberichte Heidelberger Akademie*, 1927, pp. 15-32.

[17] See Albert, *loc. cit.*[4], p. 102.

Thus for $\mathfrak{S}(\Omega_n, J_1)$ we have the basis $h_i = e_{ii} - e_{\nu+i,\nu+i}$; $e_{-\lambda_i+\lambda_j} = e_{ij} - e_{j+\nu,i+\nu}$ for $i \neq j$; $e_{-\lambda_i-\lambda_j} = e_{i,j+\nu} + e_{j,i+\nu}$, $i < j$; $e_{\lambda_i+\lambda_j} = e_{i+\nu,j} + e_{j+\nu,i}$, $i < j$; $e_{-2\lambda_i} = e_{i,i+\nu}$; $e_{2\lambda_i} = e_{i+\nu,i}$ where $i, j = 1, \cdots, \nu$. If we set $h = \Sigma h_i\lambda_i$ we obtain the following multiplication table

$$
(10)\quad
\begin{aligned}
&[h_i, h_j] = 0 \\
&[e_\alpha, h] = e_\alpha \alpha, \qquad\qquad \alpha = \pm \lambda_i \pm \lambda_j \\
&[e_\alpha, e_\beta] = \begin{cases} 0 & \text{if } \alpha + \beta \text{ is not a root} \\ e_{\alpha+\beta} N_{\alpha\beta}, \ N_{\alpha\beta} = \pm 1, \pm 2, & \text{if } \alpha + \beta \neq 0 \text{ is a root} \end{cases} \\
&[e_{-\alpha}, e_\alpha] = h_\alpha \neq 0
\end{aligned}
$$

where $h_{\lambda_i-\lambda_j} = h_i - h_j$, $h_{\lambda_i+\lambda_j} = h_i + h_j$, $h_{2\lambda_i} = h_i$ and

$$(11)\qquad h_i^p = h_i, \quad e_\alpha^p = 0.$$

For $\mathfrak{S}(\Omega_n, J_2)$ the following is a basis:

$$
\begin{array}{lll}
h_i = e_{i+1,i+1} - e_{i+\nu+1,i+\nu+1}; & e_{\lambda_i-\lambda_j} = e_{j+1,i+1} - e_{i+\nu+1,j+\nu+1}, & i \neq j; \\
e_{\lambda_i+\lambda_j} = e_{i+\nu+1,j+1} - e_{j+\nu+1,i+1}, & & i < j; \\
e_{-\lambda_i-\lambda_j} = e_{j+1,i+\nu+1} - e_{i+1,j+\nu+1}, & & i < j; \\
e_{\lambda_i} = e_{1,i+1} - e_{i+\nu+1,1}; & e_{-\lambda_i} = e_{i+1,1} - e_{1,i+\nu+1} &
\end{array}
$$

where $i, j = 1, \cdots, \nu$. The multiplication table has the same form as (10) and (11) where $h_{\lambda_i-\lambda_j} = h_i - h_j$, $h_{\lambda_i+\lambda_j} = h_i + h_j$, $h_{\lambda_i} = h_j$. For $\mathfrak{S}(\Omega_n, J_3)$ the following is a basis:

$$
\begin{array}{lll}
h_i = e_{ii} - e_{\nu+i,\nu+i}; & e_{\lambda_i-\lambda_j} = e_{ji} - e_{i+\nu,j+\nu}, & i \neq j; \\
e_{\lambda_i+\lambda_j} = e_{i+\nu,j} - e_{j+\nu,i}, & & i < j; \\
e_{-\lambda_i-\lambda_j} = e_{j,i+\nu} - e_{i,j+\nu}, & & i < j
\end{array}
$$

where $i, j = 1, \cdots, \nu$. The multiplication table is the same as before with $N_{\alpha\beta} = \pm 1$, $h_{\lambda_i-\lambda_j} = h_i - h_j$, $h_{\lambda_i+\lambda_j} = h_i + h_j$.

Now suppose $p = 2$. In this case if s is alternate in the sense that $s' = s$ and the diagonal elements of s are all 0, it is cogredient to $q = \begin{pmatrix} 0 & 1_\nu \\ 1_\nu & 0 \end{pmatrix}$. If s is symmetric and not alternate it is cogredient to 1.[18] Accordingly we have to consider two Lie algebras $\mathfrak{S}(\Omega_n, J_1)$ and $\mathfrak{S}(\Omega_n, J_2)$. If we compare with the case $p \neq 2$ we see that the present $\mathfrak{S}(\Omega_n, J_1) = \bar{\mathfrak{S}}(\Omega_n, J_1) + \mathfrak{R}$ where $\mathfrak{S}$ consists of the matrices $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ with $a_{22} = a'_{11}$, a_{12} and a_{21} alternate

---

[18] See Albert, "Symmetric and alternate matrices in an arbitrary field. I," *Transactions of the American Mathematical Society*, vol. 43 (1938), p. 392.

and $\mathfrak{K}$ has the basis $e_{i+v,i}$, $e_{i,i+v}$, $i=1,\cdots,v$. If $a \in \bar{\mathfrak{S}}$, $b \in \mathfrak{K}$ then $[a,b] \in \bar{\mathfrak{S}}$ and $b^2 \in \bar{\mathfrak{S}}$. $\bar{\mathfrak{S}}$ has the following basis:

$$\begin{array}{lll} h_i = e_{ii} + e_{v+i,v+i}; & f_{ij} = e_{ij} + e_{j+v,i+v}; & i \neq j; \\ g_{ij} = g_{ji} = e_{i,j+v} + e_{j,i+v}, & & i \neq j; \\ k_{ij} = k_{ji} = e_{i+v,j} + e_{j+v,i}, & & i \neq j \end{array}$$

and the multiplication table

$$(12)\quad \begin{array}{l} [f_{ij}, h] = f_{ij}(\lambda_i + \lambda_j),\ [g_{ij}, h] = g_{ij}(\lambda_i + \lambda_j),\ [k_{ij}, h] = k_{ij}(\lambda_i + \lambda_j) \\ [f_{ij}, f_{kl}] = \delta_{jk} f_{il} + \delta_{il} f_{kj},\ (k,l) \neq (i,j) \text{ or } (j,i) \\ [f_{ij}, f_{ji}] = h_i + h_j \\ [g_{ij}, k_{jk}] = f_{ik},\ [g_{ij}, k_{ij}] = h_i + h_j \\ [g_{ij}, f_{kj}] = g_{ik} \\ [k_{ij}, f_{jk}] = k_{ik} \end{array}$$

where i, j, k are unequal. All other products are 0 and

$$(13)\qquad h_i^2 = h_i,\quad f_{ij}^2 = 0,\quad g_{ij}^2 = 0,\quad k_{ij}^2 = 0.$$

It follows that $\bar{\mathfrak{S}}$ is the ideal generated by the elements $[c,d]$ and c^2, c, d in $\mathfrak{S}$.

The condition that $q^{-1}a'q = a$ is equivalent to $(qa)' = (qa)$ and $a \in \mathfrak{S}$ if and only if (qa) is alternate. Consider an arbitrary alternate matrix

$$b = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} \qquad b_{ij} = b_{ji},\quad b_{ii} = 0,\quad n = 2v$$

Let $\Delta = \det b$ and $B_{ij} =$ cofactor of b_{ji} in b. Then $B_{ij} = B_{ji}$ and $B_{ii} = 0$. Hence we have the equations

$$\begin{array}{ll} b_{12}B_{21} + b_{13}B_{31} + \cdots + b_{1n}B_{n1} = \Delta & \\ b_{j2}B_{21} + b_{j3}B_{31} + \cdots + b_{jn}B_{n1} = 0, & (j = 3, \cdots, n) \end{array}$$

and solving for B_{21} we obtain $B_{21}{}^2 = \Delta\Delta_2$ where Δ_2 is the determinant of an alternate matrix of $n-2$ rows and columns. We assume that it is the square of a polynomial in the b_{ij} and obtain the result that Δ has this property, say $\Delta = B^2$. The equation $B_{21}{}^2 = (BB_2)^2$ then shows that B is a factor of B_{21}.

Similarly every B_{ij} is divisible by B. If we apply this result to $b = q\lambda + qa$, λ an indeterminate, we obtain that det $(\lambda + a)$ is a square $\phi(\lambda)^2$ in $\Omega[\lambda]$ and that adj $(\lambda + a)$ is divisible by $\phi(\lambda)$. It follows in the usual manner that $\phi(a) = 0$.[19] If we use the form of $a = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and set $a_{11} = (\alpha_{ij}) = a'_{22}$ a simple computation shows that $\phi(\lambda)^2 = \lambda^n + (\alpha_{11} + \cdots + \alpha_{\nu\nu})^2\lambda^{n-2} + \cdots$. Thus $\phi(\lambda) = \lambda^\nu + (\mathrm{tr}'\, a)\lambda^{\nu-1} + \cdots$ where $\mathrm{tr}'\, a = \alpha_{11} + \cdots + \alpha_{\nu\nu}$.

Lemma 6. *If $a = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \epsilon\, \bar{\mathfrak{S}}$ where $a_{11} = (\alpha_{ij})$ then the characteristic polynomial of a has the form $[\phi(\lambda)]^2$ where $\phi(\lambda) = \lambda^\nu + (\alpha_{11} + \cdots + \alpha_{\nu\nu})\lambda^{\nu-1} + \cdots$ and $\phi(a) = 0$. This is true also when the elements of a are regarded as indeterminates.*

The second type of $\mathfrak{S}$ for $p = 2$, namely, $\mathfrak{S}(\Omega_n, J_2)$ has the basis $h_i = e_{ii}$, $e_{\lambda_i+\lambda_j} = e_{ij} + e_{ji}$ with $i < j$ and $i, j = 1, \cdots, n$. The multiplication table is

$$
\begin{aligned}
&[h_i, h_j] = 0 \\
(14)\qquad &[e_\alpha, h] = e_\alpha \alpha \\
&[e_\alpha, e_\beta] = \begin{cases} 0 \text{ if } \alpha + \beta \text{ is not a root} \\ e_{\alpha+\beta} \text{ if } \alpha + \beta \text{ is a root} \neq 0 \end{cases}
\end{aligned}
$$

and

$$
(15)\qquad h_i^2 = h_i,\; e^2_{\lambda_i+\lambda_j} = h_i + h_j.
$$

2.2. Simplicity proofs. The enveloping algebras.

Lemma 7. *If $p \neq 2$, $\mathfrak{S}(\Omega_n, J_i)$ is simple except when $n = 4$, $i = 3$. If $p = 2$, $\bar{\mathfrak{S}}(\Omega_n, J_1)$ with $\nu = n/2 > 2$ has only the proper ideals (1) of multiples of 1 and $\bar{\mathfrak{S}}'$ of elements a for which $\mathrm{tr}'\, a = 0$. $\mathfrak{S}(\Omega_n, J_2)$, $p = 2$, $n > 2$ has only the proper ideals (1) and $\mathfrak{S}'$ the elements of trace 0.*

Suppose $p \neq 2$ and let $\mathfrak{B}$ be an ideal $\neq 0$ in $\mathfrak{S}(\Omega_n, J_i)$ and $b = h_0 + \Sigma\, a_\alpha\beta_\alpha$ be an element $\neq 0$ in $\mathfrak{B}$. Choose $\bar\lambda_i$ in Ω so that the values $\bar\alpha$ of the roots α for these $\bar\lambda_i$ are all distinct and $\neq 0$. Then

$$
[\cdots[[b, \overbrace{\bar h], \bar h] \cdots \bar h}^{k}] = 0 + \Sigma\, e_\alpha\beta_\alpha\bar\alpha^k
$$

is in $\mathfrak{B}$ and since the Vandermonde determinant

[19] I am indebted to Dr. J. Williamson for the present proof.

$$\begin{vmatrix} 1 & 0 & 0 & \cdots \\ 1 & \bar{\alpha} & \bar{\alpha}^2 & \cdots \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ 1 & \bar{\epsilon} & \bar{\epsilon}^2 & \cdots \end{vmatrix} \neq 0$$

(number of rows = number of roots plus 1), h_0 and each $e_\alpha\beta_\alpha \epsilon \mathfrak{B}$. If $\beta_\alpha \neq 0$, $\mathfrak{B}$ contains e_α and if $h_0 \neq 0$ at least one $[e_\alpha, h_0] = e_\alpha \alpha_0 \neq 0$ so that we may suppose that $\mathfrak{B}$ contains an e_α. Now except in the cases $i = 1$, $n = 2$; $i = 2$, $n = 1, 3$; $i = 3$, $n = 2, 4$ it is readily verified that we may obtain any root ρ in the form $(((\alpha + \beta) + \gamma) + \delta \cdots)$ where each parenthesis $(\alpha + \beta)$, $((\alpha + \beta) + \gamma) \cdots$ is a root $\neq 0$. It follows that $[\cdots[[e_\alpha, e_\beta], e_\gamma]\cdots]$ $= e_\rho N_{\alpha\beta} N_{\beta\gamma} \cdots \neq 0$ and that $e_\rho \epsilon \mathfrak{B}$ for every ρ. Since the elements $[e_{-\alpha}, e_\alpha] = h_\alpha$ generate the subalgebra $\mathfrak{H}$ of elements h_i we obtain $\mathfrak{B} = \mathfrak{S}$. If $i = 1$, $n = 2$, $\mathfrak{B}$ contains e_α and $[-e_\alpha, e_\alpha] = h_\alpha \neq 0$ and hence also $[e_{-\alpha}, h_\alpha] = e_{-\alpha}\beta \neq 0$ so that $\mathfrak{B} = \mathfrak{S}$. If $i = 2$, $n = 1$, $\mathfrak{S} = 0$ and if $i = 2$, $n = 2$ the argument is the same as for $\mathfrak{S}(\Omega_n, J_1)$. If $i = 3$, $n = 2$, $\mathfrak{S}$ has order 1 and if $i = 3$, $n = 4$ the result does not hold.

$p = 2$. Let $\mathfrak{B}$ be an ideal in $\bar{\mathfrak{S}}$ and containing b an element $\neq 1\alpha$. If $b = h_0 + \sum_{i \neq j} f_{ij}\rho_{ij} + \sum_{i<j} g_{ij}\sigma_{ij} + \sum_{i<j} k_{ij}\tau_{ij}$ then by the above argument h_0 and $u_{ij} = f_{ij}\rho_{ij} + f_{ji}\rho_{ji} + g_{ij}\sigma_{ij} + k_{ij}\tau_{ij}$ are in $\mathfrak{B}$. If $\sigma_{ij} \neq 0$ and $k \neq i, j$ then $[[u_{ij}, k_{jk}], f_{ji}]\sigma_{ij}^{-1} = f_{jk} \epsilon \mathfrak{B}$ and similarly if $\tau_{ij} \neq 0$, $\mathfrak{B}$ contains an f_{jk}. If $\sigma_{ij} = \tau_{ij} = 0$, $\mathfrak{B}$ contains either $h_0 \neq 1\alpha$ or some $f_{ij}\rho_{ij} + f_{ji}\rho_{ji} \neq 0$. Now the Lie algebra generated by the h_i and f_{ij} is isomorphic to $\Omega_{\nu l}$. Hence by the structure of $\Omega_{\nu l}$ we obtain that $\mathfrak{B}$ contains all $\begin{pmatrix} a_{11} & 0 \\ 0 & a'_{11} \end{pmatrix}$ with $\operatorname{tr} a_{11} = 0$ and also $[g_{ij}, h_j + h_k] = g_{ij}$ and $[k_{ij}, h_j + h_k] = k_{ij}$. Thus $\mathfrak{B}$ contains all elements a of $\bar{\mathfrak{S}}$ with $\operatorname{tr}' a = 0$. If in addition $\mathfrak{B}$ contains an element not in $\bar{\mathfrak{S}}'$, $\mathfrak{B} = \bar{\mathfrak{S}}$.

In the case $\mathfrak{S}(\Omega_n, J_2)$, $p = 2$, $n > 2$ any $\mathfrak{B}$ which contains an element $b \neq 1\alpha$ contains an $e_\alpha \neq 0$ and hence all e_α and all $e_\alpha{}^2 = h_i + h_j$. It follows that $\mathfrak{B} = \mathfrak{S}(\Omega_n, J_2)'$ or $\mathfrak{B} = \mathfrak{S}(\Omega_n, J_2)$.

Corollary. *If $p = 2$ and $\nu \not\equiv 0 \pmod 2$, $\bar{\mathfrak{S}}(\Omega_n, J_1) = \bar{\mathfrak{S}}(\Omega_n, J_1)' \oplus (1)$ and $\bar{\mathfrak{S}}'$ is simple. If $\nu \equiv 0 \pmod 2$ the difference algebra $\bar{\mathfrak{S}}'/(1)$ is simple. Similarly $\mathfrak{S}(\Omega_n, J_2) = \mathfrak{S}(\Omega_n, J_2)' \oplus (1)$ when $n \not\equiv 0 \pmod 2$ and $\mathfrak{S}'$ is simple; $\mathfrak{S}'/(1)$ is simple when $n \equiv 0 \pmod 2$.*

If $\nu \equiv 0 \pmod 2$ $\bar{\mathfrak{S}}'$ does not contain (1), otherwise it does.

Remark. It should be noted that the proof of the above lemma is valid

when the algebras are regarded as ordinary Lie algebras except when $p = 2$, $i = 2$.

In the remainder of the paper we suppose that for $p \neq 2$, $i = 2$, $n \geq 3$; if $i = 3$, $n \geq 6$ and if $p = 2$, $i = 1$, $\nu \geq 3$; if $i = 2$, $n \geq 3$. With these restrictions on n we have

LEMMA 8. *If $p \neq 2$ the enveloping algebra of $\mathfrak{S}(\Omega_n, J_i)$ is Ω_n. If $p = 2$ the enveloping algebra of $\bar{\mathfrak{S}}(\Omega_n, J_1)'$ or of $\mathfrak{S}(\Omega_n, J_2)'$ is Ω_n.*

$p \neq 2$, $i = 1$. The enveloping algebra $\mathfrak{B}$ contains all matrices of the form

$$\begin{pmatrix} a & \\ & -a' \end{pmatrix}\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & ab \\ 0 & 0 \end{pmatrix} \quad \text{if} \begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix} \epsilon\, \mathfrak{B}.$$

and

$$\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}\begin{pmatrix} -a' & \\ & a \end{pmatrix} = \begin{pmatrix} 0 & ba \\ 0 & 0 \end{pmatrix}.$$

Thus for the matrices $\begin{pmatrix} 0 & b \\ 0 & 0 \end{pmatrix}$ in $\mathfrak{B}$ the set $\{b\}$ forms a two sided ideal in Ω_ν (since a is arbitrary). Since Ω_ν is a simple associative algebra every $\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}$ is in $\mathfrak{B}$ and similarly every $\begin{pmatrix} 0 & 0 \\ 0 & a \end{pmatrix}$ is in $\mathfrak{B}$. Then $\mathfrak{B}$ contains all

$$\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 0 & a \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & a \end{pmatrix}$$

and $\mathfrak{B} = \Omega_n$.

$p \neq 2$, $i = 3$. The above proof is valid.

$p \neq 2$, $i = 2$. The argument shows that $\mathfrak{B}$ contains all matrices of the form

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & a_{11} & a_{12} \\ 0 & a_{21} & a_{22} \end{pmatrix}$$

where a_{ij} are arbitrary in Ω_ν. $\mathfrak{B}$ contains also

$$(e_{12} - e_{\nu+1,1})(e_{1,\nu+1} - e_{21}) = -e_{11} - e_{\nu+1,\nu+1}$$

and hence

$$e_{11} \text{ and } e_{1,i+1} = e_{11}(e_{1,i+1} - e_{i+\nu+1,1}), \quad e_{i+\nu+1,1} = -(e_{1,i+1} - e_{i+\nu+1,1})e_{11}$$

and similarly $e_{i+1,1}$, $e_{1,i+\nu+1}$. Hence $\mathfrak{B} = \Omega_n$.

$p=2,\ i=1$. The enveloping algebra of matrices of the form $\begin{pmatrix} a & 0 \\ 0 & a' \end{pmatrix}$ with $\operatorname{tr} a = 0$ includes the matrices

$$\begin{pmatrix} a_1 & \\ & a_2 \end{pmatrix}, \qquad \begin{pmatrix} a_2 & \\ & a_1 \end{pmatrix} \qquad ,$$

where a_1 is arbitrary and a_2 depends on a_1. Then by the argument in the first case $\mathfrak{B} = \Omega_n$.

$p=2,\ i=2$. $\mathfrak{B}$ contains $(e_{ii}+e_{jj})(e_{ik}+e_{ki}) = e_{ik}$ if i, j, k are unequal and $e_{ii} = e_{ik}e_{ki}$. Hence $\mathfrak{B} = \Omega_n$.

COROLLARY. *The sets of matrices of Lemma 8 are irreducible sets.*

2.3. Restricted derivations.

LEMMA 9. *If $p \neq 2$ the restricted derivations of $\mathfrak{S}(\Omega_n, J_i)$ are all inner. The same holds for $p=2$ and $i=2$. The restricted derivations of $\bar{\mathfrak{S}}(\Omega_n, J_1)$ are induced by inner derivations of Ω_{nl}.*

$p \neq 2$ or $p=2,\ i=2$. Choose $\bar{\lambda}_i$ in Ω so that the values $\Sigma\, m_i\bar{\lambda}_i$ of the p^n linear forms $\Sigma\, m_i\lambda_i$ are distinct. Then the minimum p-polynomial of

$\bar{h} = \Sigma\, h_i\bar{\lambda}_i$ is $\Pi(\lambda - \Sigma\, m_i\bar{\lambda}_i) = \lambda^{p^\nu} + \cdots + \lambda\beta_0$, $\beta_0 = \prod_{\neq 0} (\Sigma\, m_i\bar{\lambda}_i) \neq 0$.

Thus

$$\bar{h}^{p^\nu} + \cdots + \bar{h}\beta_0 = 0.$$

If $p=2,\ i=2$ the same result holds with n in place of ν. Suppose D is a restricted derivation. Then

$$(\bar{h}D)(\bar{H}^{p^\nu-1} + \cdots + \beta_0) = (\bar{h}D) \prod_{\neq 0} (\bar{H} - \Sigma\, m_i\lambda_i) = 0$$

where $\bar{H}$ is the mapping $x \to [x, \bar{h}]$. Hence $\bar{h}D = \Sigma\, e_\alpha\rho_\alpha$. If $\sigma_\alpha = -\bar{\alpha}^{-1}\rho_\alpha$, $[\bar{h}, \Sigma\, e_\alpha\sigma_\alpha] = \bar{h}D$. By subtracting the inner derivation $x \to [x, \Sigma\, e_\alpha\sigma_\alpha]$ from D we obtain the restricted derivation E such that $\bar{h}E = 0$. It follows that $\bar{h}^pE = \bar{h}^{p^2}E = \cdots = 0$ and since $\bar{h}, \bar{h}^p, \cdots$ form a basis for $\mathfrak{H}$, $hE = 0$ for all h in $\mathfrak{H}$. Since $[e_\alpha, \bar{h}] = e_\alpha\bar{\alpha}$, $[e_\alpha E, h] = e_\alpha E$ and $e_\alpha E = e_\alpha\xi_\alpha$. Since $[e_{-\alpha}, e_\alpha] = h_\alpha \neq 0$, $[e_{-\alpha}\xi_{-\alpha}, e_\alpha] + [e_{-\alpha}, e_\alpha\xi_\alpha] = 0$ and $\xi_{-\alpha} = -\xi_\alpha$. Similarly $[e_\alpha, e_\beta] = e_{\alpha+\beta}N_{\alpha+\beta}$ yields $\xi_{\alpha+\beta} = \xi_\alpha + \xi_\beta$ if $\alpha+\beta \neq 0$ is a root. By a fundamental set of roots (f. s.) we shall mean a set of roots $\alpha_1, \alpha_2, \cdots, \alpha_\nu$ (or $\alpha_1, \alpha_2, \cdots, \alpha_{n-1}$ if $p=i=2$) which are linearly independent and such that any root ρ may be obtained as $((\pm\alpha_{i_1} \pm \alpha_{i_2}) \pm \alpha_{i_3}) \pm \cdots$ where each

parenthesis is a root $\neq 0$. For $p \neq 2$, $i = 1$, $\lambda_1 - \lambda_2, \cdots, \lambda_1 - \lambda_\nu$, $\lambda_1 + \lambda_2$ is a f. s.; $p \neq 2$, $i = 2$, $\lambda_1, \lambda_2 \cdots \lambda_\nu$ is a f. s.; $p \neq 2$, $i = 3$, $\lambda_1 - \lambda_2, \cdots, \lambda_1 - \lambda_\nu$, $\lambda_1 + \lambda_2$ is a f. s.; $p = 2$, $i = 1$, $\lambda_1 + \lambda_2, \cdots, \lambda_1 + \lambda_\nu$; $p = i = 2$, $\lambda_1 + \lambda_2, \cdots, \lambda_1 + \lambda_n$ is a f. s. Now let $\alpha_1, \alpha_2, \cdots, \alpha_\nu$ $(\alpha_1, \cdots, \alpha_{n-1})$ be a f. s. By the linear independence we may solve the equation $\alpha_i = \xi_{\alpha_i}$ for $\lambda_i = \lambda_{i0}$ in Ω. Then if $h_0 = \Sigma h_i \lambda_{i0}$ and H_0 is the inner derivation $x \to [x, h_0]$ we have $h(E - H_0) = 0$, $e_{\alpha_i}(E - H_0) = 0$. Hence by the above $e_{-\alpha_i}(E - H_0) = 0$ and if $e_\rho = [\cdot \cdot \cdot [e_\alpha, e_\beta], e_\gamma \cdot \cdot \cdot] N_{\alpha\beta}^{-1} N_{\beta\gamma}^{-1} \cdots$ where $\alpha, \beta, \cdots = \pm \alpha_i$ we have $e_\rho(E - H_0) = 0$. Thus $E = H_0$ and D is inner.

$p = 2$, $i = 1$. If D is a restricted derivation the above argument shows that we may subtract an inner derivation from D to obtain E such that $hE = 0$, $h \in \mathfrak{H}$. Then

$$f_{ij}E = f_{ij}\alpha_{ij} + f_{ji}\beta_{ji} + g_{ij}\gamma_{ij} + k_{ij}\epsilon_{ij}.$$

Since $f_{ij}^2 E = [f_{ij}E, f_{ij}] = 0$, $\beta_{ji} = 0$. Now $k_i = e_{i,i+\nu}$ and $k_{i+\nu} = e_{i+\nu,i}$ are in $\mathfrak{S}$ and $[f_{i1}, k_{i+\nu}] = k_{1i}$, $[f_{1i}, k_i] = g_{1i}$ if $i > 1$ while all other products with f_{j1}, f_{1j} are 0. We note also that $[h, k_i] = [h, k_{i+\nu}] = 0$. Hence we may subtract the derivation $x \to [x, k]$ where $k = \Sigma k_i \gamma_{1i} + \Sigma k_{\nu+i}\epsilon_{i1}$ from E and obtain a restricted derivation F such that $hF = 0$, $f_{1i}F = f_{1i}\alpha_{1i} + k_{1i}\epsilon_{1i}$, $f_{i1}F = f_{i1}\alpha_{i1} + g_{1i}\gamma_{i1}$. Next we subtract a suitable derivation of the form $x \to [x, h_0]$, h_0 in $\mathfrak{H}$, to obtain G such that $hG = 0$, $f_{1i}G = k_{1i}\epsilon_{1i}$. Since $[f_{i1}, f_{1i}] = h_1 + h_i$ it follows that $f_{i1}G = g_{1i}\gamma_{i1}$. Since $[f_{i1}, k_1] = g_{1i}$, $[f_{1i}, k_{1+\nu}] = k_{1i}$ we can obtain by subtraction a K such that $hK = f_{21}K = f_{12}K = 0$, $f_{i1}K = g_{1i}\gamma_{i1}$, $f_{1i}K = k_{1i}\epsilon_{1i}$ for $i > 2$. Operating with K on $f_{21} = [[f_{21}, f_{1i}], f_{i1}]$ gives $\gamma_{i1} = \epsilon_{1i} = 0$ so that $f_{i1}K = f_{1i}K = 0$ and hence $f_{ij}K = 0$. Then

$$g_{ij}K = g_{ij}\rho_{ij} + f_{ij}\sigma_{ij} + f_{ji}\sigma_{ji} + k_{ij}\tau_{ij}, \quad \rho_{ij} = \rho_{ji}, \quad \tau_{ij} = \tau_{ji}$$

and since $[g_{ij}K, g_{ij}] = 0$, $\tau_{ij} = 0$. Let $k \neq i, j$. Then $[g_{ij}, f_{kj}] = g_{ik}$ and hence

$$g_{ik}\rho_{ik} + f_{ik}\sigma_{ik} + f_{ki}\sigma_{ki} = g_{ik}\rho_{ij} + f_{ki}\sigma_{ji}$$

and $\sigma_{ik} = 0$, $\rho_{ik} = \rho_{ij}$ or $\rho_{ij} = \rho$. In the same way we find that $k_{ij}K = k_{ij}\rho'$ and since $[g_{ij}, k_{ij}] = h_i + h_j$, $\rho = \rho'$. Choose α and β so that $\alpha + \beta = \rho$ and define $g = (\Sigma e_{ii})\alpha + (\Sigma e_{i+\nu,i+\nu})\beta$. Then $xK = [x, g]$.

2.4. Automorphisms. Let S be an automorphism in $\mathfrak{S}(\Omega_n, J_i)$ with $p \neq 2$ and $\{\Lambda\}$ the weights of h in the representation $x \to x^S$. Since this

representation is $(1-1)$ there are ν independent weights. Since $(h^S)^{J_i} = -h^S$, if Λ is a weight so is $-\Lambda$. Hence if $i = 1, 3$ the weights are $\Lambda_1, \cdots, \Lambda_\nu, -\Lambda_1, \cdots, -\Lambda_\nu$ and if $i = 2$, $\Lambda_1, \cdots, \Lambda_\nu, 0, -\Lambda_1, \cdots, -\Lambda_\nu$ where $\Lambda_1, \cdots, \Lambda_\nu$ are linearly independent. Thus any subset of unequal weights not containing 0 is linearly independent if and only if it does not contain a weight together with its negative.

Evidently if α is a root, $\Lambda + \alpha, \Lambda, \Lambda - \alpha$ are linearly dependent. Hence if these are weights and $\Lambda \neq 0$ we have either $\Lambda + \alpha = -\Lambda$, $-\Lambda = \Lambda - \alpha$ or $\Lambda + \alpha = 0$, $\Lambda - \alpha = 0$. In the first two cases we have $\Lambda = \pm \alpha/2$ and $3\alpha/2, \alpha/2, -(\alpha/2)$ or $\alpha/2, -(\alpha/2), -(3\alpha/2)$ are weights.[20] Then $-(3\alpha/2)$ or $3\alpha/2$ is a weight and since $\alpha/2 \neq 3\alpha/2, \neq -(3\alpha/2)$ this is impossible. The last two possibilities $\Lambda = \alpha, -\alpha$ can occur only when $i = 2$, for otherwise no weight is 0.

Suppose Λ is a weight $\neq \pm \alpha, \neq 0$ and $\Lambda + \alpha$ is not a weight but $\Lambda - \alpha$ is. If $\Lambda - 2\alpha$ is a weight, $\Lambda, \Lambda - \alpha, \Lambda - 2\alpha$ are dependent. Hence either $-\Lambda = \Lambda - \alpha$, $-\Lambda + \alpha = \Lambda - 2\alpha$, $-\Lambda = \Lambda - 2\alpha$ or $\Lambda = 2\alpha$. The first two may be excluded as before, the third since $\Lambda \neq \alpha$. If $\Lambda = 2\alpha$ then the weights are $2\alpha, \alpha, 0$. Then $-2\alpha, -\alpha$ are weights and this is impossible unless $p = 3$ and then we have a contradiction with $\Lambda \neq -\alpha$. Let x_0 be a vector such that $x_0 h^S = x_0 \Lambda$. Since $\Lambda + \alpha$ is not a weight $x_0 e_\alpha{}^S = 0$ and since one of $\Lambda - \alpha, \Lambda - 2\alpha$ is not a weight, either $x_0 e_{-\alpha}{}^S = 0$ or $x_0 e_{-\alpha}{}^S \neq 0$ and $x_0 (e_{-\alpha}{}^S)^2 = 0$. By Lemma 1 this implies $\Lambda_\alpha = 0$ or $\Lambda_\alpha = \alpha_\alpha / 2$.

$i = 1, 3$. Let $\Lambda = \Sigma\, m_i \lambda_i$ be a weight, α a root. Either $\Lambda + \alpha$ or $\Lambda - \alpha$ is not a weight and we may suppose we have the first case. If $\alpha = \lambda_i - \lambda_j$, $h_\alpha = h_i - h_j$ and $\alpha_\alpha = 2$. Hence $\Lambda_\alpha = m_i - m_j = 0, 1$. If $\Lambda - \alpha$ were a weight we would obtain $m_i - m_j = 0, -1$. If $\Lambda_\alpha = 1$,

$$\Lambda - \alpha = \Lambda - (m_i - m_j)\alpha = \sum_{k \neq i, j} m_k \lambda_k + m_j \lambda_i + m_i \lambda_j$$

is a weight and similarly if $m_i - m_j = -1$, $\Lambda + \alpha = \Lambda - (m_i - m_j)\alpha$ is a weight. If $\alpha = \lambda_i + \lambda_j$, $h_\alpha = h_i + h_j$, $\alpha_\alpha = 2$. Here we obtain $m_i + m_j = 0$, ± 1 and if $m_i + m_j = 1$, $\Lambda - (m_i + m_j)\alpha = \Sigma\, m_k \lambda_k - m_j \lambda_i - m_i \lambda_j$. Then by the above $\Sigma\, m_k \lambda_k - m_i \lambda_i - m_j \lambda_j$ is a weight. Combining these results we obtain the following condition on the weights: The set of weights is invariant under permutation of the λ's and under change of sign of an even number of λ's. Either all $m_i = \pm \frac{1}{2}$ or one $m_i = \pm 1$ and all others are 0. If $i = 1$ we may use the root $\alpha = 2\lambda_i$ and obtain $m_i = 0$, or ± 1. Hence in this case

[20] Since we are assuming that $\Lambda + \alpha, \Lambda, \Lambda - \alpha$ are $\neq 0$, $p \neq 3$.

the weights are $\pm\lambda_i$. If $i=3$ we must consider the possibility $m_i=\pm\frac{1}{2}$. By changing signs and permuting we obtain the weight

$$\tfrac{1}{2}(\lambda_1+\cdots+\lambda_{\nu-2})-\tfrac{1}{2}(\lambda_{\nu-1}+\lambda_\nu) \quad \text{or} \quad \tfrac{1}{2}(\lambda_1+\cdots+\lambda_{\nu-1})-\tfrac{1}{2}\lambda_\nu.$$

The first case is excluded if $\nu\geq 5$ for then we obtain by permuting the λ's more than ν linearly independent weights. In the second case if $\nu\geq 5$ we obtain ν independent weights by permuting and the additional weight $\frac{1}{2}(\lambda_1+\cdots+\lambda_{\nu-3})-\frac{1}{2}(\lambda_{\nu-2}+\lambda_{\nu-1}+\lambda_\nu)$ independent of these. If $\nu=3$ we would obtain the weights $\pm\frac{1}{2}\lambda_1\pm\frac{1}{2}\lambda_2\pm\frac{1}{2}\lambda_3$ and this is impossible since there are only six weights. If $\nu=4$ we can not exclude the following two additional possibilities $\Lambda_j=\frac{1}{2}(\sum_1^4\lambda_i)-\lambda_j$ for $j=1,\cdots,4$ and

$$\Lambda_1=\tfrac{1}{2}\lambda_1+\tfrac{1}{2}\lambda_2+\tfrac{1}{2}\lambda_3+\tfrac{1}{2}\lambda_4, \qquad \Lambda_2=\tfrac{1}{2}\lambda_1+\tfrac{1}{2}\lambda_2-\tfrac{1}{2}\lambda_3-\tfrac{1}{2}\lambda_4,$$
$$\Lambda_3=\tfrac{1}{2}\lambda_1-\tfrac{1}{2}\lambda_2+\tfrac{1}{2}\lambda_3-\tfrac{1}{2}\lambda_4, \qquad \Lambda_4=\tfrac{1}{2}\lambda_1-\tfrac{1}{2}\lambda_2-\tfrac{1}{2}\lambda_3+\tfrac{1}{2}\lambda_4.$$

Thus unless $i=3$, $\nu=4$ the weights of h in the representation $x\rightarrow x^S$ are $\pm\lambda_i$.

$i=2$. If Λ is a weight $\neq 0$, and $\neq$ any root the argument is similar to that for $i=3$ and shows that $m_i-m_j=0,\pm 1$, $m_i+m_j=0,\pm 1$ and that the forms obtained from Λ by permuting the m's are weights. Since $\alpha=\lambda_i$ is a root, $h_{\lambda_i}=h_i$, and $\Lambda_\alpha=m_i$, $\alpha_\alpha=1$. Hence we obtain $m_i=0,\pm\frac{1}{2}$ and in the latter case $\Lambda-\alpha$ is a weight and is obtained from Λ by changing the sign of one m. It follows that all $\frac{1}{2}(\Sigma\pm\lambda_j)$ are weights and this is impossible if $\nu>2$. Thus every $\Lambda\neq 0$ is a root. If $\Lambda=\lambda_i+\lambda_j$, $\alpha=\lambda_j$ then $\Lambda_\alpha=1$ and $\Lambda-\alpha=\lambda_i=\Lambda'$ is a weight. Then if $\beta=\lambda_i-\lambda_j$, $\Lambda'-\beta=\lambda_j$ is a weight. This is impossible as is $\Lambda=\pm\lambda_i\pm\lambda_j$. Hence the weights here are $\pm\lambda_i$ also.

LEMMA 10. *If $p\neq 2$ and S is an automorphism of $\mathfrak{S}(\Omega_n,J_i)$ where if $i=3$, $\nu\neq 4$ then there is a matrix q independent of λ such that $h^S=q^{-1}hq$.*

This is equivalent to the result established that the weights of h^S are $\pm\lambda_i$ if $i=1,3$ and $0,\pm\lambda_i$ if $i=2$.

Now suppose $p=2$, $i=1$. If $a=\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a'_{11}\end{pmatrix}$ where $a'_{12}=a_{12}$, $a'_{21}=a_{21}$ with diagonal elements all 0 is in $\bar{\mathfrak{S}}(\Omega_n,J_1)$, we defined $\mathrm{tr}'\,a=\mathrm{tr}\,a_{11}$. Then

$$a^2=\begin{pmatrix} a_{11}^2+a_{12}a_{21} & a_{11}a_{12}+a_{12}a'_{11} \\ a_{21}a_{11}+a'_{11}a_{21} & a'_{11}{}^2+a_{21}a_{12}\end{pmatrix}$$

and if $a_{12} = (\alpha_{ij})$, $a_{21} = (\beta_{ij})$ we have $\operatorname{tr} a_{12}a_{21} = \Sigma\, \alpha_{ij}\beta_{ji} = 0$ by the conditions on a_{12}, a_{21}. Hence $\operatorname{tr}' a^2 = (\operatorname{tr}' a)^2$ and if and only if $\nu \equiv 0 \pmod 2$ the correspondence $a \to a + \operatorname{tr}' a$ is $(1-1)$ and hence an automorphism in $\bar{\mathfrak{S}}$.

Let S be any automorphism in $\bar{\mathfrak{S}}$. Let Λ be a weight in the representation $x \to x^S$ and x_0 a vector such that $x_0 h^S = x_0 \Lambda$. If either $x_0 f_{ij}{}^S$ or $x_0 f_{ji}{}^S \neq 0$ we obtain $\Lambda + (\lambda_i + \lambda_j)$ as a weight also and if $x_0 f_{ij}{}^S = x_0 f_{ji}{}^S = 0$, $x_0(h_i{}^S + h_j{}^S) = 0$ and hence $m_i = m_j$ in $\Lambda = \Sigma\, m_i \lambda_i$. Thus the weights are invariant under permutation of the λ's. Since the characteristic polynomial of h^S is a perfect square the multiplicity of each Λ is even. If $\Lambda = \lambda_{k_1} + \cdots + \lambda_{k_\mu}$ is a weight we obtain, by permuting, $\binom{\nu}{\mu}$ distinct weights. Since there are ν linearly independent weights we may suppose that $0 < \mu < \nu$. By the assumption $\nu \geqq 3$ we obtain either $\mu = 1$ or $\mu = \nu - 1$. In the former case the weights are λ_i and each has multiplicity two. If $\mu = \nu - 1$ the weights are $\Sigma\, \lambda_j + \lambda_i$, $i = 1, \cdots, \nu$. Then $1^S = \Sigma\, h_i{}^S = 1(\nu - 1) \neq 0$ and if T denotes the automorphism $a \to a + \operatorname{tr}' a$ we see that ST has the weights λ_i each with multiplicity 2.

$p = 2$, $i = 2$. In this case $a \to a + \operatorname{tr} a$ is an automorphism if $n \equiv 0 \pmod 2$. Let S be an automorphism and x_0 a vector $\neq 0$ such that $x_0 h^S = x_0 \Lambda$. Either $x_0 e_\alpha{}^S = 0$, $\alpha = \lambda_i + \lambda_j$ in which case $x_0(e_\alpha{}^S)^2 = x_0(h_i + h_j) = 0$, $m_i = m_j$ in $\Lambda = \Sigma\, m_i\lambda_i$ or $\Lambda + \alpha$ is a weight. Thus the weights are invariant under permutation and as above they are either λ_i, $i = 1, \cdots, n$ or $\Sigma\, \lambda_j + \lambda_i$ and in the latter case $n - 1 \not\equiv 0 \pmod 2$. We may then form ST, $a^T = a + \operatorname{tr} a$ and obtain an automorphism with weights λ_i.

LEMMA 11. *If $p = 2$ and S is an automorphism of $\bar{\mathfrak{S}}(\Omega_n, J_1)$ $(\mathfrak{S}(\Omega_n, J_2))$ and $\nu \not\equiv 0 \pmod 2$ $(n \not\equiv 0 \pmod 2)$ then there exists a matrix q independent of λ such that $h^S = q^{-1}hq$. If $\nu \equiv 0 \pmod 2$ $(n \equiv 0 \pmod 2)$ either this holds or the automorphism ST where $a^T = a + \operatorname{tr}' a$ $(a^T = a + \operatorname{tr} a)$ has this property.*

LEMMA 12. *The matrix q in Lemma 10 or 11 may be normalized so that $qq^{J_i} = 1 = q^{J_i}q$.*

$p \neq 2$. Since $h^S = q^{-1}hq$, $-h^S = h^{SJ_i} = -q^{J_i}h(q^{J_i})^{-1}$. Hence $(qq^{J_i})h = h(qq^{J_i})$. It follows that $r = qq^{J_i}$ is a diagonal matrix. Since $r^{J_i} = r$ a simple computation shows that

$$r = \{\rho_1, \cdots, \rho_\nu, \rho_1, \cdots, \rho_\nu\}, \quad = \{\rho, \rho_1, \cdots, \rho_\nu, \rho_1, \cdots, \rho_\nu\},$$
$$= \{\rho_1, \cdots, \rho_\nu, \rho_1, \cdots, \rho_\nu\}$$

according as $i = 1, 2, 3$. It follows that we can find a diagonal matrix k such that $kk^{J_i} = r^{-1}$. The matrix $u = kq$ has the desired property.

$p = 2$, $i = 2$. In this case $r = qq^{J_2}$ is diagonal. Hence as above we may solve $h^2 = -r^{-1}$ and obtain $u = hq$ of the desired type.

$p = 2$, $i = 1$. Here $r = qq^{J_1}$ has the form

$$\begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}$$

where the M_{ij} are diagonal. If

$$q = \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} \qquad r = qq^{J_1} = \begin{pmatrix} q_{11}q'_{22} + q_{12}q'_{21} & q_{11}q'_{12} + q_{12}q'_{11} \\ q_{21}q'_{22} + q_{22}q'_{21} & q_{21}q'_{12} + q_{22}q'_{11} \end{pmatrix}$$

and hence $M_{11} = M'_{22}$ and $M_{12} = b + b' = 0$, $M_{21} = c + c' = 0$. In this case too we may solve the equation $kk^{J_1} = r^{-1}$ to obtain a k commutative with h. The matrix kq satisfies the condition.

If U denotes the inverse $a \to uau^{-1}$ of the mapping determined by Lemma 10 or 11, the correspondence $\bar{S} = SU$ (or STU) is an automorphism in $\mathfrak{S}(\Omega_n, J_i)$ such that $h^{\bar{S}} = h$. From $[e_a, h] = e_a\alpha$ we obtain when $p \neq 2$ and when $p = i = 2$ $e_a{}^{\bar{S}} = e_a\xi_a$. If $p \neq 2$, $[e_{-a}, e_a] = h_a$ and hence $\xi_{-a} = \xi_a{}^{-1}$. Let $\alpha_1, \alpha_2, \cdots$ be the fundamental set of roots determined above. These are for $p \neq 2$ and $i = 1$, $\lambda_1 - \lambda_2, \cdots, \lambda_1 - \lambda_\nu, \lambda_1 + \lambda_2$; for $p \neq 2$, $i = 2$, λ_j and for $p \neq 2$, $i = 3$, $\lambda_1 - \lambda_2, \cdots, \lambda_1 - \lambda_\nu, \lambda_1 + \lambda_2$. If $i = 1, 3$ we solve the equations $\bar{\lambda}_1\bar{\lambda}_j{}^{-1} = \xi_{\lambda_1 - \lambda_j}$, $\bar{\lambda}_1\bar{\lambda}_2 = \xi_{\lambda_1+\lambda_2}$; if $i = 2$ solve $\bar{\lambda}_i = \xi_{\lambda_i}$. Then for $i = 1, 2, 3$, respectively, the matrices

$$g = \{\bar{\lambda}_1, \cdots, \bar{\lambda}_\nu, \bar{\lambda}_1{}^{-1}, \cdots, \bar{\lambda}_\nu{}^{-1}\}, \quad = \{1, \bar{\lambda}_1, \cdots, \bar{\lambda}_\nu, \bar{\lambda}_1{}^{-1}, \cdots, \bar{\lambda}_\nu{}^{-1}\}$$
$$= \{\bar{\lambda}_1, \cdots, \bar{\lambda}_\nu, \bar{\lambda}_1{}^{-1}, \cdots, \bar{\lambda}_\nu{}^{-1}\}$$

satisfy $g^{J_i}g = 1 = gg^{J_i}$ and $a^{\bar{S}} = g^{-1}ag$ for $a = h, e_{a_i}$. Hence the latter equation holds for e_{-a_i} and hence for all e_a. Thus $a^{\bar{S}} = g^{-1}ag$ for all a in $\mathfrak{S}$.

$p = 2$, $i = 2$. Here $e_a{}^2 = h_i + h_j$ and hence $\xi_a{}^2 = 1$, $\xi_a = 1$ and $\bar{S} = 1$. We consider finally the case $p = 2$, $i = 1$. Here

$$f_{ji}{}^{\bar{S}} = f_{ji}\alpha_{ji} + f_{ij}\beta_{ij} + g_{ij}\gamma_{ij} + k_{ij}\epsilon_{ij} \tag{16}$$

where the condition $(f_{j1}{}^{\bar{S}})^2 = 0$ yields $\alpha_{j1}\beta_{1j} = \gamma_{1j}\epsilon_{1j}$. If g is J_1-orthogonal we have seen that $x \to g^{-1}xg$ is an automorphism in $\mathfrak{S}$. Since $\bar{\bar{\mathfrak{S}}}$ is a characteristic subalgebra $g^{-1}xg \in \bar{\bar{\mathfrak{S}}}$ if x does. Now the matrix

$$d = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}$$

where

$$d_{11} = \{\mu_1, \cdots, \mu_\nu\},\ d_{12} = \{\rho_1, \cdots, \rho_\nu\},\ d_{21} = \{\sigma_1, \cdots, \sigma_\nu\},\ d_{22} = \{\tau_1, \cdots, \tau_\nu\}$$

is J_1-orthogonal if and only if $d_{11}d_{22} + d_{12}d_{21} = 1$. The automorphism $a \to \bar{a} = d^{-1}ad = d^{J_1}ad$ sends the elements of $\mathfrak{H}$ into themselves and

$$\begin{aligned}
\bar{f}_{j1} &= f_{j1}\mu_1\tau_j + f_{1j}\rho_1\sigma_j + g_{1j}\rho_1\tau_j + k_{1j}\mu_1\sigma_j \\
\bar{f}_{1j} &= f_{1j}\tau_1\mu_j + f_{j1}\rho_j\sigma_1 + g_{1j}\tau_1\rho_j + k_{1j}\sigma_1\mu_j \\
\bar{g}_{1j} &= f_{ij}\tau_1\sigma_j + f_{j1}\tau_j\sigma_1 + g_{1j}\tau_1\tau_j + k_{1j}\sigma_i\sigma_j \\
\bar{k}_{1j} &= f_{1j}\rho_1\mu_j + f_{ji}\mu_1\rho_j + g_{1j}\rho_1\rho_j + k_{1j}\mu_1\mu_j.
\end{aligned}$$

Since $\alpha_{21}\beta_{12} = \gamma_{12}\epsilon_{12}$ we may solve the equations $\mu_1\tau_2 = \alpha_{21}$, $\rho_1\sigma_2 = \beta_{12}$, $\rho_1\tau_2 = \gamma_{12}$, $\mu_1\sigma_2 = \epsilon_{12}$. For this solution we can not have simultaneously $\mu_1 = 0$, $\rho_1 = 0$ or $\sigma_2 = 0$, $\tau_2 = 0$ since $f_{21}^{\bar{S}} \neq 0$. Hence we may determine the other parameters so that $d^{J_1}d = 1$. The product of $\bar{S}$ and the inverse of $a \to \bar{a}$ is an automorphism $\bar{S}_1$ such that $f_{21}^{\bar{S}_1} = f_{21}$. Since $[f_{j1}, f_{21}] = 0$, $[f_{j1}^{\bar{S}_1}, f_{21}] = 0$ and if we use the notation (16) we obtain $\beta_{1j} = \gamma_{1j} = 0$, $j = 3, \cdots$. Since we do not have $\alpha_{j1} = \epsilon_{1j} = 0$ we may find a d of the above type for which $\mu_1 = \tau_1 = \mu_2 = \tau_2 = 1$, $\rho_1 = \sigma_1 = \rho_2 = \sigma_2 = 0$ and $\tau_j = \alpha_{j1}$, $j = 3, \cdots, \sigma_j = \epsilon_{1j}$. For the corresponding automorphism we have $f_{j1} = f_{j1}^{\bar{S}_1}$, $j = 2, \cdots$. Then by multiplying $\bar{S}_1$ by the inverse of a suitable transformation $a \to \bar{a}$ we obtain $\bar{S}_2$ such that $f_{j1}^{\bar{S}_2} = f_{j1}$, $j = 2, \cdots, \nu$. Now

$$f_{1j}^{\bar{S}_2} = f_{1j}\alpha_{1j} + f_{j1}\beta_{j1} + g_{j1}\gamma_{j1} + k_{j1}\epsilon_{j1}$$

and if $k \neq j$ the condition $[[f_{k1}, f_{1j}], f_{j1}] = f_{k1}$ implies $\alpha_{1j} = 1$ and $(f_{1j}^{\bar{S}_2})^2 = 0$, implies $\beta_{j1} = \gamma_{j1}\epsilon_{j1}$. Choose $\sigma_1 = \epsilon_{21}$, $\rho_2 = \gamma_{21}$, the other σ's and ρ's $= 0$, $\mu_i = \tau_i = 1$. The automorphism determined satisfies $\bar{h} = h$, $\bar{f}_{j1} = f_{j1}$, $\bar{f}_{12} = f_{12}^{\bar{S}_2}$. We then obtain an $\bar{S}_3$ such that $h^{\bar{S}_3} = h$, $f_{j1}^{\bar{S}_3} = f_{j1}$, $f_{12}^{\bar{S}_3} = f_{12}$. Then by $[f_{1j}, f_{12}] = 0$ we obtain $\beta_{j1} = \epsilon_{j1} = 0$, $f_{1j}^{\bar{S}_3} = f_{1j} + g_{j1}\gamma_{j1}$, $j = 3, \cdots, \nu$. Let $\rho_j = \gamma_{j1}$, $j = 3, \cdots$ all the other ρ's $= 0$, $\sigma_i = 0$, $\mu_i = \tau_i = 1$. We then obtain $\bar{h} = h$, $\bar{f}_{j1} = f_{j1}$, $\bar{f}_{1j} = f_{1j}^{\bar{S}_3}$ and therefore the product $\bar{S}_4$ of $\bar{S}_3$ and the inverse of a suitable transformation $a \to \bar{a}$ satisfies $h^{\bar{S}_4} = h$, $f_{j1}^{\bar{S}_4} = f_{j1}$, $f_{1j}^{\bar{S}_4} = f_{1j}$. Then $f_{ij}^{\bar{S}_4} = f_{ij}$ and

$$g_{1j}^{\bar{S}_4} = f_{1j}\mu_{1j} + f_{j1}\rho_{j1} + g_{1j}\sigma_{j1} + k_{1j}\tau_{j1}.$$

Using $g_{1k} = [g_{1j}, f_{kj}]$ we obtain $\mu_{1j} = \tau_{j1} = 0$, $\rho_{j1} = \rho_{k1} = \rho$, $\sigma_{j1} = \sigma_{k1} = \sigma$ and from $g_{1k} = [[g_{1j}, f_{k1}], f_{1j}]$ we obtain $\rho = 0$. Similarly we obtain $k_{1j}^{\bar{S}_4} = k_{1j}\zeta$. Since $[g_{1j}, k_{1j}] = h_1 + h_j$, $\sigma\zeta = 1$. Finally we set $\rho_i = \sigma_i = 0$, $\mu_i^2 = \zeta$, $\tau_i^2 = \sigma$ and obtain $\bar{h} = h^{\bar{S}_4}$, $\bar{f}_{ij} = f_{ij}^{\bar{S}_4}$, $\bar{g}_{1j} = g_{1j}^{\bar{S}_4}$, $\bar{k}_{1j} = k_{1j}^{\bar{S}_4}$. Since

$g_{jk} = [g_{1j}, f_{k1}]$, $g_{jk}^{\bar{S}_i} = \bar{g}_{jk}$ and similarly $\bar{k}_{jk} = k_{jk}^{\bar{S}_i}$. Hence $\bar{a} = a^{\bar{S}_i}$. We have therefore proved the following

LEMMA 13. *If $p \neq 2$ the automorphisms of $\mathfrak{S}(\Omega_n, J_i)$ all have the form $a \to g^{-1}ag$ where $g^{J_i}g = 1\gamma$ if $i = 1$, ν arbitrary; $i = 2$, $\nu \geq 3$; $i = 3$, $\nu \geq 3$, $\neq 4$. The same result holds when $p = 2$ for $\bar{\mathfrak{S}}(\Omega_n, J_1)$ with $\nu \not\equiv 0 \pmod 2$ $\nu \geq 3$ and for $\mathfrak{S}(\Omega_n, J_2)$ with $n \not\equiv 0 \pmod 2$ $n \geq 3$. In the case $\bar{\mathfrak{S}}(\Omega_n, J_1)$, $\nu \equiv 0 \pmod 2$ $\nu \geq 3$ we have in addition the automorphisms $a \to g^{-1}ag + \mathrm{tr}' a$ and for $\mathfrak{S}(\Omega_n, J_2)$, $n \equiv 0 \pmod 2$ $n \geq 3$ the remaining automorphisms are $a \to g^{-1}ag + \mathrm{tr}\, a$.*

2.5. Isomorphisms. The dimensionality of $\mathfrak{S}(\Omega_n, J_i)$, $i = 1, 2, 3$, is respectively $\nu(2\nu + 1)$, $\nu(2\nu + 1)$ and $\nu(2\nu - 1)$ for $p \neq 2$. Hence the only isomorphisms that can occur are between $\mathfrak{S}(\Omega_{2\nu}, J_1)$ and $\mathfrak{S}(\Omega_{2\nu+1}, J_2)$. Now suppose S is an isomorphism between $\mathfrak{S}(\Omega_{2\nu+1}, J_2)$ and $\mathfrak{S}(\Omega_{2\nu}, J_1)$. Consider the weights of h in the representation $a \to a^S$ of $\mathfrak{S}(\Omega_{2\nu+1}, J_2)$. Since h^S is skew, if Λ is a weight so is $-\Lambda$ and since h_i^S are linearly independent there are ν linearly independent weights and hence these are $\Lambda_1, \cdots, \Lambda_\nu$ with the remaining weights $-\Lambda_1, \cdots, -\Lambda_\nu$. If α is a root of $\mathfrak{S}(\Omega_{2\nu+1}, J_2)$ and Λ is a weight either $\Lambda + \alpha$ or $\Lambda - \alpha$ is not a weight. It follows that $\Lambda_\alpha = 0$, ± 1 for $\alpha = \lambda_i - \lambda_j$ or $\lambda_i + \lambda_j$ and hence if $\Lambda = \Sigma\, m_i \lambda_i$, $m_i \pm m_j = 0, \pm 1$. We have also as in the last section that the weights are invariant under permutation of the λ's and arbitrary changes of sign of the m's. Hence if $\nu \geq 3$ the weights are $\pm \lambda_i$. Thus h^S is similar to the h given by our particular basis for $\mathfrak{S}(\Omega_{2\nu}, J_1)$. It follows by the proof of Lemma 12 that we may suppose that the similarity is given by a J_1-orthogonal element. Hence by combining S with a suitable automorphism we obtain an isomorphism T such that $h^T = h$ in $\mathfrak{S}(\Omega_{2\nu}, J_1)$. Then $[e_{\lambda_i}{}^T, h^T] = e_{\lambda_i}\lambda_i$ and this is impossible since λ_i is not a root in $\mathfrak{S}(\Omega_{2\nu}, J_1)$.

If $\nu = 1$, the basis of $\mathfrak{S}(\Omega_2, J_1)$ is e_+, e_-, h such that $e_+^p = 0$, $e_-^p = 0$, $h^p = h$ and if we replace these respectively by $e_+/\sqrt{2}$, $e_-/\sqrt{2}$, $h/2$ we obtain a basis whose multiplication table is that of the e's and h of $\mathfrak{S}(\Omega_3, J_2)$. If $\nu = 2$ it is readily verified that the following is an isomorphism between $\mathfrak{S}(\Omega_4, J_1)$ and $\mathfrak{S}(\Omega_5, J_2)$

$$\begin{array}{ll}
h_1 \to h_1 + h_2 & h_2 \to h_1 - h_2 \\
e_{\lambda_1-\lambda_2} \to -e_{\lambda_2}\sqrt{2} & e_{\lambda_2-\lambda_1} \to -e_{-\lambda_2}\sqrt{2} \\
e_{\lambda_1+\lambda_2} \to e_{\lambda_1}\sqrt{2} & e_{-\lambda_1-\lambda_2} \to e_{-\lambda_1}\sqrt{2} \\
e_{2\lambda_1} \to e_{\lambda_1+\lambda_2} & e_{-2\lambda_1} \to e_{-\lambda_1-\lambda_2} \\
e_{2\lambda_2} \to e_{\lambda_1-\lambda_2} & e_{-2\lambda_2} \to e_{\lambda_2-\lambda_1}
\end{array}$$

If $p = 2$ the algebras $\bar{\mathfrak{S}}(\Omega_{2\nu}, J_1)$ are not isomorphic to any $\mathfrak{S}(\Omega_n, J_2)$ since the former have outer derivations while the derivations of the latter are all inner.

LEMMA 13. *The only isomorphic pairs of algebras in the set* $\mathfrak{S}(\Omega_n, J_i)$, $\bar{\mathfrak{S}}(\Omega_n, J_1)$ *are* $\mathfrak{S}(\Omega_{n2}, J_1)$, $\mathfrak{S}(\Omega_3, J_2)$ *and* $\mathfrak{S}(\Omega_4, J_1)$, $\mathfrak{S}(\Omega_5, J_2)$.

2. 6. Restricted Lie algebras of types *B, C, D*. Let $\mathfrak{A}$ be a normal simple associative algebra over Φ and suppose that $\mathfrak{A}$ has an involution J of the first kind.[9] The set $\mathfrak{S}(\mathfrak{A}, J)$ of J-skew elements is a restricted Lie algebra relative to $a + b$, $a\alpha$, $[a, b] = ab - ba$ and a^p. If Ω is the algebraic closure of Φ, $a_1, a_2, \cdots, a_{n^2}$ a basis for $\mathfrak{A}$ over Φ then $\mathfrak{A}_\Omega = \Omega_n$ and the correspondence $\Sigma\, a_i\omega_i \to \Sigma\, a_i^J\omega_i$ is an involution J in Ω_n. The skew elements relative to J form the extension $\mathfrak{S}(\mathfrak{A}, J)_\Omega$. On the other hand we may take this set to be one of the sets $\mathfrak{S}(\Omega_n, J_i)$ considered above. Following the notation introduced by Cartan in the characteristic 0 case we shall say that $\mathfrak{S}(\mathfrak{A}, J)$ is of type C, B, D according as $i = 1, 2, 3$. Thus if $p \neq 2$, $\mathfrak{S}(\mathfrak{A}, J)$ is a simple restricted Lie algebra except when its type is D and $n = 4$. If $p = 2$ and $\mathfrak{S}(\mathfrak{A}, J)_\Omega = \mathfrak{S}(\Omega_n, J_1)$ the algebra $\bar{\mathfrak{S}}(\mathfrak{A}, J)$ generated by the elements a^2 and $[a, b]$ extends to $\bar{\mathfrak{S}}(\mathfrak{A}, J)_\Omega = \bar{\mathfrak{S}}(\Omega_n, J_1)$ and hence has order $\nu(2\nu - 1)$, $\nu = n/2$.[21] Similarly $\bar{\mathfrak{S}}(\mathfrak{A}, J)'$, the subalgebra of $\bar{\mathfrak{S}}$ generated by $[a, b], a, b$ in $\bar{\mathfrak{S}}$, extends to $\bar{\mathfrak{S}}(\Omega_n, J_1)'$. It follows that if ν is odd and ≥ 3, $\bar{\mathfrak{S}}(\mathfrak{A}, J)'$ is simple and if ν is even and ≥ 4, $\bar{\mathfrak{S}}(\mathfrak{A}, J)'/(1)$ is simple. For $p = 2$ and $\mathfrak{S}(\mathfrak{A}, J)_\Omega = \mathfrak{S}(\Omega_n, J_2)$ and $\mathfrak{S}(\mathfrak{A}, J)'$ the algebra generated by $[a, b]$ extends to $\mathfrak{S}(\Omega_n, J_2)'$. Hence $\mathfrak{S}(\mathfrak{A}, J)'$ is simple if n is odd and ≥ 3 and $\mathfrak{S}(\mathfrak{A}, J)'/(1)$ is simple when n is even and ≥ 4. We note also that $\bar{\mathfrak{S}}(\Omega_n, J_2) = \mathfrak{S}(\Omega_n, J_2)$ and hence that $\bar{\mathfrak{S}}(\mathfrak{A}, J) = \mathfrak{S}(\mathfrak{A}, J)$.

THEOREM 9. *If $\mathfrak{A}$ is a normal simple associative algebra of order n^2 with an involution J of the first kind, then $\mathfrak{S}(\mathfrak{A}, J)$, the set of J-skew elements, is a simple restricted Lie algebra when $p \neq 2$ except when $\mathfrak{S}$ is of type D and $n = 4$. If $p = 2$ either $\bar{\mathfrak{S}}(\mathfrak{A}, J)'$ or $\bar{\mathfrak{S}}(\mathfrak{A}, J)'/(1)$ is simple when $\mathfrak{S}(\mathfrak{A}, J)_\Omega = \mathfrak{S}(\Omega_n, J_1)$ and $\nu = n/2 \geq 3$ and either $\mathfrak{S}(\mathfrak{A}, J)$ or $\mathfrak{S}(\mathfrak{A}, J)'/(1)$ is simple when $\mathfrak{S}(\mathfrak{A}, J)_\Omega = \mathfrak{S}(\Omega_n, J_2)$ and $n \geq 3$.*

Since the coefficients of the characteristic polynomials of the elements of $\mathfrak{A}$ in the representation in Ω_n all belong to Φ it follows that if $p = 2$ and

---

[21] In general, for any restricted Lie algebra $\mathfrak{L}$ we denote the sub-algebra generated by the elements $[a, b]$ and a by $\bar{\mathfrak{L}}$. It follows that $\bar{\mathfrak{L}}_P = \overline{(\mathfrak{L}_P)}$.

$a \in \mathfrak{S}(\mathfrak{A}, J)'$ where $\mathfrak{S}(\mathfrak{A}, J)_\Omega = \mathfrak{S}(\Omega_n, J_1)$ then $\mathrm{tr}' a \in \Phi$. Hence the correspondence $a \to a + \mathrm{tr}' a$ is an automorphism if $\nu \equiv 0 \pmod 2$. Similarly for the other type of $\mathfrak{S}(\mathfrak{A}, J)$ with $p = 2$ we have the automorphisms $a \to a + \mathrm{tr}\, a$ for $n \equiv 0 \pmod 2$.

In the remainder of Part II we make the following restrictions on the orders: In Theorems 10 and 13, $\nu \geq 1$, ≥ 3, ≥ 3 but $\neq 4$ according as the type is C, B or D. Otherwise the restrictions are as in Lemma 13. Now if $\mathfrak{S}(\mathfrak{A}, J)$ and $\mathfrak{S}(\mathfrak{B}, K)$ are isomorphic they are of the same type. If $p \neq 2$ and $a \to a^G$ is an isomorphism between two restricted Lie algebras $\mathfrak{S}(\mathfrak{A}, J)$ and $\mathfrak{S}(\mathfrak{B}, K)$ then we may regard these as subsets of the same $\mathfrak{S}(\Omega_n, J_i)$ such that the elements of $\mathfrak{S}(\mathfrak{A}, J)$ are $\Sigma a_i \phi_i$ uniquely, those of $\mathfrak{S}(\mathfrak{B}, K)$ are $\Sigma a_i^G \phi_i$ and those of $\mathfrak{S}(\Omega_n, J_i)$ are representable uniquely either as $\Sigma a_i \omega_i$ or $\Sigma a_i^G \omega_i$. Thus the correspondence $\Sigma a_i \omega_i \to \Sigma a_i^G \omega_i$ is an automorphism in $\mathfrak{S}(\Omega_n, J_i)$. Hence by Lemma 13 we obtain

THEOREM 10. *A necessary and sufficient condition that the restricted Lie algebras $\mathfrak{S}(\mathfrak{A}, J)$ and $\mathfrak{S}(\mathfrak{B}, K)$ be isomorphic is that $\mathfrak{A}$ and $\mathfrak{B}$ be isomorphic and J and K be cogredient.*

THEOREM 11. *Any automorphism G in $\mathfrak{S}(\mathfrak{A}, J)$ has the form $a^G = g^{-1}ag$ where g is a J-orthogonal element.*

THEOREM 12. *Any restricted derivation in $\mathfrak{S}(\mathfrak{A}, J)$ is inner.*

If $p = 2$ we cannot have an isomorphism between $\bar{\mathfrak{S}}(\mathfrak{A}, J)$ and $\bar{\mathfrak{S}}(\mathfrak{B}, K)$ unless either $\bar{\mathfrak{S}}(\mathfrak{A}, J)_\Omega = \bar{\mathfrak{S}}(\mathfrak{B}, K) = \bar{\mathfrak{S}}(\Omega_n, J_1)$ or $\bar{\mathfrak{S}}(\mathfrak{A}, J)_\Omega = \bar{\mathfrak{S}}(\mathfrak{B}, K)_\Omega = \bar{\mathfrak{S}}(\Omega_n, J_2)$. In the former case an isomorphism between the two algebras becomes an automorphism in $\bar{\mathfrak{S}}(\Omega_n, J_1)$ and may have the form $a \to g^{-1}ag + \mathrm{tr}' a = b$. But then $\mathrm{tr}' a = \mathrm{tr}' b$ since $\nu \equiv 0 \pmod 2$ and hence by combining the isomorphism with the automorphism $b \to b + \mathrm{tr}' b$ we obtain the isomorphism $a \to g^{-1}ag$. In a similar fashion we may treat the other case. We obtain the following results:

THEOREM 13. *If $p = 2$, $\bar{\mathfrak{S}}(\mathfrak{A}, J)$ and $\bar{\mathfrak{S}}(\mathfrak{B}, K)$ are isomorphic if and only if $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic and J and K are cogredient.*

THEOREM 14. *If $p = 2$ the following are the automorphisms of $\bar{\mathfrak{S}}(\mathfrak{A}, J)$*

$a \to g^{-1}ag$, *g J-orthogonal if* $\mathfrak{S}_\Omega = \mathfrak{S}(\Omega_{2\nu}, J_1)$ *and* $\nu \not\equiv 0 \pmod 2$ *or if* $\mathfrak{S}_\Omega = \mathfrak{S}(\Omega_n, J_2)$ *and* $n \not\equiv 0 \pmod 2$

$a \to g^{-1}ag$, $a \to g^{-1}ag + \mathrm{tr}' a$ *if* $\mathfrak{S}_\Omega = \mathfrak{S}(\Omega_{2\nu}, J_1)$ *and* $\nu \equiv 0 \pmod 2$.
$a \to g^{-1}ag$, $a \to g^{-1}ag + \mathrm{tr}\, a$ *if* $\mathfrak{S}_\Omega = \mathfrak{S}(\Omega_n, J_2)$ *and* $n \equiv 0 \pmod 2$.

THEOREM 15. *The restricted derivations of* $\bar{\bar{\mathfrak{S}}}(\mathfrak{A}, J)$ *are inner if* $p = 2$ *and* $\mathfrak{S}(\mathfrak{A}, J)_\Omega = \mathfrak{S}(\Omega_n, J_1)$. *For the other case* $\bar{\bar{\mathfrak{S}}}(\mathfrak{A}, J)$ *any restricted derivation is induced by a derivation of* $\mathfrak{A}$.

If $\mathfrak{S}_\Omega = \mathfrak{S}(\Omega_n, J_1)$ and $\nu \not\equiv 0 \pmod 2$ we have $\bar{\bar{\mathfrak{S}}}(\mathfrak{A}, J) = \bar{\bar{\mathfrak{S}}}(\mathfrak{A}, J)' \oplus (1)$. Any automorphism (derivation) in $\bar{\bar{\mathfrak{S}}}'$ is induced by an automorphism (derivation) of $\bar{\bar{\mathfrak{S}}}$. Hence these have the form $a \to g^{-1}ag\,(a \to [a, d], d$ in $\mathfrak{A})$. If $\mathfrak{S}_\Omega = \mathfrak{S}(\Omega_n, J_2)$ and $n \not\equiv 0 \pmod 2$, $\bar{\bar{\mathfrak{S}}}(\mathfrak{A}, J) = \bar{\bar{\mathfrak{S}}}(\mathfrak{A}, J)' \oplus (1)$. The automorphisms have the form $a \to g^{-1}ag$ and the derivations $a \to [a, d]$, d in $\bar{\bar{\mathfrak{S}}}'$.

We remark finally that Theorem 9 for $p \neq 2$ or $p = 2$ and $\mathfrak{S}_\Omega = \mathfrak{S}(\Omega_n, J_1)$, Theorems 10, 11 and the remarks of the last paragraph hold when these algebras are regarded as ordinary Lie algebras rather than restricted Lie algebras. For these algebras have no centers.

2.7. Completeness. We wish to determine the restricted Lie algebras $\mathfrak{L}$ such that $\mathfrak{L}_\Omega$ is one of the simple algebras $\mathfrak{S}(\Omega_n, J_i)$, $i = 1, 2, 3$ for $p \neq 2$, or $\bar{\bar{\mathfrak{S}}}(\Omega_n, J_1)'$ for $p = 2$, $\nu \not\equiv 0 \pmod 2$, $\bar{\bar{\mathfrak{S}}}(\Omega_n, J_2)'$, $p = 2$, $n \not\equiv 0 \pmod 2$. If $a_1, a_2, \cdots, a_n$ is a basis for $\mathfrak{L}$ over Φ we may suppose that the elements $\Sigma\, a_j\phi_j$ of $\mathfrak{L}$ are subsets of the corresponding $\mathfrak{S}$ or $\bar{\bar{\mathfrak{S}}}'$ such that the elements of $\mathfrak{S}$ or $\bar{\bar{\mathfrak{S}}}'$ are representable uniquely in the form $\Sigma\, a_j\omega_j$. The elements of the matrices a_j and hence of $\Sigma\, a_j\phi_j$ determine a finite extension Γ of Φ. Thus the e_α, h basis of $\mathfrak{S}$ or $\bar{\bar{\mathfrak{S}}}'$ is expressible in terms of the a_j with coefficients in Γ and so we may replace Ω by Γ: $\mathfrak{L}_\Gamma = \mathfrak{S}(\Gamma_n, J_i)$ or $\bar{\bar{\mathfrak{S}}}(\Gamma_n, J_i)'$. Let $[a_j, a_k] = \Sigma\, a_p\alpha_{pjk}$, $a_j^p = \Sigma\, a_q\beta_{qj}$, $\alpha, \beta \in \Phi$ be the multiplication table of the a's.

We suppose first that $\Gamma = \Phi(x_1, x_2, \cdots, x_r, y)$ is a direct product of $\Phi(x_1, \cdots, x_r)$, where $x_i^p = \xi_i \in \Phi$ and $\Phi(y)$ is a normal separable field over Φ. Let $\mathfrak{g} = (i, s, t, \cdots)$ be the Galois group of Γ over $\Phi(x_1, \cdots, x_r)$ and d a derivation in Γ such that the d-constants are the elements of $\Phi(y)$.[16] Thus the elements of Φ are the only ones which satisfy $\alpha^d = 0$, $\alpha^s = \alpha$ for all s in $\mathfrak{g}$. If $c = (\gamma_{ij})$ is any element of Γ_n the correspondence $c \to c^d = (\gamma_{ij}{}^d)$ is a derivation in Γ_n over Φ and $c \to c^s = (\gamma_{ij}{}^s)$ is an automorphism in Γ_n over Φ. Since the matrices s_i which define the involution in Γ_n have elements in Φ these correspondences induce, respectively, a restricted derivation and an automorphism in $\mathfrak{L}_\Gamma = \mathfrak{S}(\Gamma_n, J_i)$ or $\bar{\bar{\mathfrak{S}}}(\Gamma_n, J_i)'$ over Φ. Since the constants of multiplication, $\alpha, \beta \in \Phi$ the correspondence $\Sigma\, a_j\gamma_j \to \Sigma\, a_j{}^d\gamma_j = aD$ is a restricted derivation in $\mathfrak{L}_\Gamma$ over Γ and hence is induced by a derivation D in the associative algebra Γ_n over Γ. Similarly $a = \Sigma\, a_j\gamma_j \to \Sigma\, a_j{}^s\gamma_j = a^S$ is induced by an automorphism S in Γ_n over Γ. Let $D_1 = d - D$, $S_1 = sS^{-1}$.

These are respectively derivations and automorphisms in Γ_n over Φ. Hence the elements b such that $bD_1 = 0$, $bS_1 = b$ for all S_1 form an algebra $\mathfrak{A}$ over Φ. Since $(\Sigma a_j\gamma_j)D_1 = \Sigma a_j\gamma_j^d$, $(\Sigma a_j\gamma_j)S_1 = \Sigma a_j\gamma_j^s$, $\mathfrak{L}_\Gamma \wedge \mathfrak{A} = \mathfrak{L}$. Since the enveloping algebra of $\mathfrak{L}_\Gamma$ is Γ_n we may obtain a basis for Γ_n of the form $a_1, a_2, \cdots, a_{n^2}$ where a_i is a product of the a_j's, $j = 1, \cdots, m$. Each $a_i \in \mathfrak{A}$. Hence if $(\Sigma a_i\gamma_i)D_1 = \Sigma a_i\gamma_i^d = 0$ and $(\Sigma a_i\gamma_i)S_1 = \Sigma a_i\gamma_i$, $\gamma_i = \phi_i \in \Phi$. $\mathfrak{A}$ therefore consists of the set $\Sigma a_i\phi_i$ and $\mathfrak{A}_\Gamma = \Gamma_n$. Thus $\mathfrak{A}$ is normal simple. Since the $a_i = a_{k_1} \cdots a_{k_t}$, $a_i^{J_i} = (-1)^t a_{k_t} \cdots a_{k_1} \in \mathfrak{A}$ and J_i induces an involution in $\mathfrak{A}$. Since the a_j form a basis for the J_i-skew elements of Γ_n or for $\bar{\mathfrak{S}}(\Gamma_n, J_i)'$ they also form a basis for those sets in $\mathfrak{A}$. Thus $\mathfrak{L} = \mathfrak{S}(\mathfrak{A}, J_i)$ or $\bar{\mathfrak{S}}(\mathfrak{A}, J_i)'$.

Now let Γ be any finite extension and Δ the maximal separable subfield of Γ. There is a chain of fields $\Delta_0 = \Delta, \Delta_1, \cdots, \Delta_u = \Gamma$ where

$$\Delta_i = \Delta_{i-1}(x_1, \cdots, x_{r_i}) > \Delta_{i-1}, \qquad x_j^p = \xi_j \in \Delta_{i-1}.$$

Now

$$(\mathfrak{L}_{\Delta_{u-1}})_{\Delta_u} = \mathfrak{S}(\Delta_{u,n}, J_i) \quad \text{or} \quad \bar{\mathfrak{S}}(\Delta_{u,n}, J_i)'.$$

Hence by what has been proved $\mathfrak{L}_{\Delta_{u-1}} = \mathfrak{S}(\mathfrak{A}_u, J)$ or $\bar{\mathfrak{S}}(\mathfrak{A}_u, J)'$. Let $\bar{\Gamma}$ be a separable extension of Δ_{u-1}. It is known that if $\bar{\Delta}$ is the maximal separable over Φ subfield of $\bar{\Gamma}$ then there is a chain of subfields of the above type such that $\bar{\Gamma} = \bar{\Delta}_{u-1}$.[17] If E is a splitting field for $\mathfrak{A}_u$ over Δ_u, $\mathfrak{L}_E = \mathfrak{S}(E_n, J)$ or $\bar{\mathfrak{S}}(E_n, J)'$ and the involution J has the form $a \to s^{-1}a's$ where $s' = \pm s$. If $p \neq 2$ and $s' = -s$ or $p = 2$ and $s' = s$ is alternate we may replace s by the cogredient matrix $\begin{pmatrix} 0 & 1_v \\ -1_v & 0 \end{pmatrix}$ or $\begin{pmatrix} 0 & 1_v \\ 1_v & 0 \end{pmatrix}$. Thus $\mathfrak{L}_E = \mathfrak{S}(E_n, J_1)$ or $\bar{\mathfrak{S}}(E_n, J_1)'$. If $p \neq 2$ and $s' = s$ we adjoin certain square roots to E and obtain s cogredient to s_2 or s_3 of (9). Hence by first choosing E to be separable we may obtain a separable $\bar{\Gamma} = \bar{\Delta}_{u-1}$ over Δ_{u-1} such that $\mathfrak{L}_{\bar{\Gamma}} = \mathfrak{S}(\bar{\Gamma}_n, J_i)$ or $\mathfrak{L}_{\bar{\Gamma}} = \bar{\mathfrak{S}}(\bar{\Gamma}_n, J_1)'$. If $p = 2$ and $s' = s$ there is, by a lemma we shall prove below, a $\bar{\Gamma}$ of the form $\bar{\Delta}_{u-1}$ over $\Phi(b_1, b_2 \cdots b_t)$, b_i^p in Φ such that $\mathfrak{L}_{\bar{\Gamma}} = \mathfrak{S}(\Gamma_n, J_2)$. By continuing this process we obtain an extension Γ^* of Φ which is either separable or is a separable extension of a $\Phi(b_1, \cdots, b_v)$, $b_i^p \in \Phi$ such that $\mathfrak{L}_{\Gamma^*} = \mathfrak{S}(\Gamma^*_n, J_i)$ or $\bar{\mathfrak{S}}(\Gamma^*_n, J_i)'$. If Γ^* is separable we replace it by a separable normal field. In the other case Γ^* is a direct product of $\Phi(b_k)$ and a separable field over Φ and the latter may be replaced by a separable normal field. In either case we obtain from $\mathfrak{L}_{\Gamma^*} = \mathfrak{S}$ or $\bar{\mathfrak{S}}'$ that $\mathfrak{L} = \mathfrak{S}(\mathfrak{A}, J)$ or $\bar{\mathfrak{S}}(\mathfrak{A}, J)'$ where $\mathfrak{A}$ is normal simple over Φ.

LEMMA 14. *If $\mathfrak{L}$ is a restricted Lie algebra over Φ of characteristic 2 such that there exists an extension* P *of* Φ *with the property* $\mathfrak{L}_{\mathrm{P}} = \mathfrak{S}(\mathrm{P}_n, J)'$ *where* n *is odd and* J *is the involution* $a \to s^{-1}a's$, $s' = s$, *not alternate, then there is an extension* Σ *of the form* $\Sigma = \Phi(b_1, \cdots, b_t, \mathrm{P})$ *where* $b_j^2 = \beta_j \in \Phi$ *such that* $\mathfrak{L}_{\Sigma} = \mathfrak{S}(\Sigma_n, J_2)'$.

Without loss of generality we may suppose that

$$s = \{\alpha_1, \alpha_2, \cdots, \alpha_n\}.$$

Then $\mathfrak{S}(\mathrm{P}_n, J)$ has the basis $h_i = e_{ii}$ and $e_{\lambda_i+\lambda_j} = e_{ij} + e_{ji}\alpha_i\alpha_j^{-1}$ with $i < j$ and the multiplication table

$$\begin{aligned} &[e_\alpha, h] = e_\alpha \alpha \text{ if } h = \Sigma h_i \lambda_i \\ &[e_\beta, e_\alpha] = \begin{cases} 0 \text{ if } \alpha + \beta \text{ is not a root } \lambda_i + \lambda_j \\ e_{\beta+\alpha} N_{\alpha\beta} \neq 0 \text{ if } \alpha + \beta \text{ is a root} \end{cases} \\ &h_i^2 = h_i, \qquad e_{\lambda_i+\lambda_j}{}^2 = (h_i + h_j)\alpha_i\alpha_j^{-1}. \end{aligned}$$

Since $[[e_\beta, e_\alpha], e_\alpha] = [e_\beta, e_\alpha{}^2] = e_\beta \alpha_i \alpha_j^{-1}$ if $\alpha = \lambda_i + \lambda_j$ and $\alpha + \beta$ is a root $(\neq 0)$ we have $N_{\alpha\beta} N_{\alpha+\beta,\alpha} = \alpha_i \alpha_j^{-1}$. The matrix of $\Sigma h_i \lambda_i + \Sigma e_\alpha \mu_\alpha$ in the adjoint representation has the form

$$\begin{pmatrix} 0 & X \\ 0 & M \end{pmatrix} \tag{16}$$

relative to the basis h_i and e_α where the 0-block in the upper corner has n rows and columns. M has α's down the diagonal and $\mu_\alpha N_{\alpha\beta}$ in the intersection of the " βth " row and $\alpha + \beta$-column if $\alpha + \beta$ is a root, otherwise the elements of M are 0. Thus if $\alpha + \beta$ is a root we also have $(\alpha + \beta) + \alpha$ as a root and hence $\mu_\alpha N_{\alpha+\beta,\alpha}$ is in the $(\alpha + \beta, \beta)$-place. We note also that there are $2(n-2)$ roots β such that $\alpha + \beta$ is a root for a fixed α. It follows that the characteristic polynomial of (16) is

$$\lambda^n(\lambda^m + \lambda^{m-2}\phi_2 + \cdots), \qquad (m = n(n-1)/2)$$

where

$$\begin{aligned} \phi_2(\lambda_i, \mu_\alpha) &= \Sigma\, (\lambda_1 + \lambda_2)(\lambda_1 + \lambda_3) + (n-2)\, \Sigma\, \mu^2_{\lambda_i+\lambda_j} \alpha_i \alpha_j^{-1} \\ &= \Sigma\, (\lambda_1 + \lambda_2)(\lambda_1 + \lambda_3) + \Sigma\, \mu_{ij}{}^2 \alpha_i \alpha_j^{-1}. \end{aligned}$$

We understand by the first summation the sum of all products of pairs of distinct form $\lambda_i + \lambda_j$, $\lambda_k + \lambda_l$ and we have set $\mu_{\lambda_i+\lambda_j} = \mu_{ij}$. Since $\mathfrak{S}(\mathrm{P}_n, J) = \mathfrak{S}(\mathrm{P}_n, J)' \oplus (1)$ the characteristic polynomial of $\Sigma h_i \lambda_i + \Sigma e_\alpha \mu_\alpha$ in $\mathfrak{S}'$ is $\lambda^{n-1}(\lambda^m + \lambda^{m-2}\phi_2 + \cdots)$. Thus we obtain the characteristic polynomial in the

adjoint representation of $\mathfrak{S}'$ by setting $\Sigma \lambda_i = 0$ or $\lambda_n = \lambda_1 + \cdots + \lambda_{n-1}$. Then ϕ_2 becomes

$$\phi'_2 = \sum_{i=1}^{n-1} \lambda_i^2 + \sum_{i<j=1}^{n-1} \lambda_i \lambda_j + \Sigma \mu_{ij}^2 \alpha_i \alpha_j^{-1}$$

For

$$\Sigma(\lambda_1 + \lambda_2)(\lambda_1 + \lambda_3) = \tfrac{1}{2}(n-2)(n-1) \sum_{1}^{n} \lambda_i^2 + n(n-2) \sum_{i<j=1}^{n} \lambda_i \lambda_j$$

and this becomes $\sum_{1}^{n-1} \lambda_i^2 + \sum_{i<j=1}^{n-1} \lambda_i \lambda_j$ when we set $\lambda_n = \lambda_1 + \cdots + \lambda_{n-1}$. We may apply Albert's theory of quadratic forms mod 2.[22] The matrix of the form in the λ's is

$$A = \begin{pmatrix} 1 & 0 & \cdot & \cdot & 0 \\ 1 & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & 0 \\ 1 & \cdot & \cdot & 1 & 1 \end{pmatrix}.$$

Since $A + A' = 1 + B$ where B is the matrix all of whose elements are 1 we have $(1 + B)^2 = 1 + B^2$ and since $(n-1)$ is even $B^2 = 0$. Thus $A + A'$ is non-singular. Hence the form in the λ's may be reduced to $\Sigma \lambda'^2_i \gamma'_i + \lambda'_1 \lambda'_{q+1} + \lambda'_2 \lambda'_{q+2} + \cdots \lambda'_q \lambda'_{2q}$ where $q = (n-1)/2$ and a reduction of ϕ'_2 is obtained by adding $\Sigma \mu_{ij}^2 \alpha_i \alpha_j^{-1}$.

Now we have assumed that $\mathfrak{S}(\mathrm{P}_n, J)' = \mathfrak{L}_{\mathrm{P}}$ where $\mathfrak{L}$ is a Lie algebra over Φ. This means that $\mathfrak{S}'$ has a basis a_j with a multiplication table in Φ and if we use this basis to determine the characteristic polynomial of the general element $\Sigma a_j \nu_j$ we obtain for the second coefficient a quadratic form $\psi_2(\nu)$ with coefficients in Φ equivalent to ϕ'_2. Since $n-1$ is an invariant ψ'_2 may be reduced to $\sum_{1}^{n-1} \nu'^2_i \gamma_i + \nu'_1 \nu'_{q+1} + \cdots + \nu'_q \nu'_{2q} + \sum_{n}^{m} \nu'^2_j \beta_j$, $\gamma \in \Phi$. It follows that $\Sigma \mu_{ij}^2 \alpha_i \alpha_j^{-1}$ and $\sum_{n}^{m} \nu'^2_j \beta_j$ are equivalent. If we adjoin $\sqrt{\beta_j}$ to Φ we obtain $\Sigma \mu_{ij}^2 \alpha_i \alpha_j^{-1}$, equivalent to $\Sigma \nu'^2_j$. Hence each $\alpha_i \alpha_j^{-1}$ is a square in $\Sigma = \Phi(\mathrm{P}, b_1, \cdots, b_t)$, $b_k^2 = \beta_k$, $t = n - m$ and the matrix $s\alpha_1^{-1}$ is cogredient to 1. Hence $\mathfrak{L}_\Sigma = \mathfrak{S}(\Sigma_n, J_2)'$. We have therefore proved the lemma and with it the

THEOREM 16. *If $\mathfrak{L}$ is a restricted Lie algebra over Φ such that $\mathfrak{L}_\Omega$ is one of the simple algebras $\mathfrak{S}(\Omega_n, J_i)$, $i = 1, 2, 3$, for $p \neq 2$ or $\bar{\mathfrak{S}}(\Omega_n, J_1)'$ for $p = 2$ and* $\nu \not\equiv 0 \pmod 2$ *or $\mathfrak{S}(\Omega_n, J_2)'$ for $p = 2$ and* $n \not\equiv 0 \pmod 2$ *then $\mathfrak{L}$ is isomorphic to an $\mathfrak{S}(\mathfrak{A}, J)$ to an $\bar{\mathfrak{S}}(\mathfrak{A}, J)'$ or to an $\mathfrak{S}(\mathfrak{A}, J)'$ respectively.*

[22] *Loc. cit.*[18], p. 400.

III. APPLICATIONS.

3. 1. Isomorphisms between algebras of different types. Suppose that $p \neq 2$ and that there is an isomorphism between $\mathfrak{S}(\Omega_m, J_i)$ and Ω'_{nl} $n \not\equiv 0 \pmod p$. This isomorphism maps the maximum commutative subalgebra $\mathfrak{H}$ of $\mathfrak{S}$ into a maximal commutative subalgebra $\bar{\mathfrak{H}}$ of Ω'_{nl}. By modifying the isomorphism we may suppose that $\bar{\mathfrak{H}}$ is contained in the subalgebra with basis $e_{ii} - e_{nn}$, $i = 1, \cdots, n-1$. In order to be maximal $\bar{\mathfrak{H}}$ must coincide with this algebra. Thus in the case $\mathfrak{S}(\Omega_m, J_1)$, $m = 2\mu$, we therefore have $n - 1 = \mu$ and hence the orders $n^2 - 1 = \mu(2\mu + 1)$ so that $\mu = 1$. For $\mathfrak{S}(\Omega_m, J_2)$, $m = 2\mu + 1$, we obtain the one possibility $\mu = 1$ and for $\mathfrak{S}(\Omega_m, J_3)$, $m = 2\mu$, the only possibility is $\mu = 3$. In these particular cases it is readily seen that isomorphisms exist between $\mathfrak{S}(\Omega_2, J_1)$ and Ω'_{2l} between $\mathfrak{S}(\Omega_3, J_2)$ and Ω'_{2l} and between $\mathfrak{S}(\Omega_6, J_3)$ and Ω'_{4l}.

Consider next $\Omega'_{nl}/(1)$ with $n \equiv 0 \pmod p$. Any restricted derivation D in Ω_{nl} is inner and induces a restricted derivation in $\Omega'_{nl}/(1)$. In order that D be 0 in the difference algebra it is necessary that $[a, b]D = 1\rho \in \Omega$ for all a and b in Ω_{nl}. It follows that if $u \in \Omega'_{nl}$ then $u^p D = 0$. Using $u = e_{ii} - e_{nn}$ we obtain that $xD = [x, h_0]$, h_0 in $\mathfrak{H}$. Then $[e_a, h_0] = e_a \alpha_0 = \rho_a$ implies that $h_0 = 1\sigma$, $D = 0$. Hence the algebra $\Omega'_{nl}/(1)$ has outer derivations and can not be isomorphic to any $\mathfrak{S}(\Omega_m, J_i)$. A similar discussion for the case $p = 2$ yields no isomorphic pairs.

If we use the completeness theorems we see that for $p \neq 2$ any restricted Lie algebra $\mathfrak{S}(\mathfrak{A}, J)$ such that $\mathfrak{S}(\mathfrak{A}, J)_\Omega = \mathfrak{S}(\Omega_6, J_3)$ is isomorphic to a $\mathfrak{B}'_l$, $\mathfrak{B}$ normal simple of order 16 or to an $\mathfrak{S}(\mathfrak{B}, K)'$ where $\mathfrak{B}$ is an involutorial algebra of order 16 over the center $\mathrm{P} = \Phi(q)$ and K is an involution of the second kind. Conversely any Lie algebra of one of the last two types is isomorphic to an $\mathfrak{S}(\mathfrak{A}, J)$. Similarly any $\mathfrak{S}(\mathfrak{A}, J)$ with $\mathfrak{S}(\mathfrak{A}, J)_\Omega = \mathfrak{S}(\Omega_3, J_2)$ or $\mathfrak{S}(\Omega_5, J_2)$ is isomorphic to an $\mathfrak{S}(\mathfrak{B}, K)$ with $\mathfrak{S}(\mathfrak{B}, K)_\Omega = \mathfrak{S}(\Omega_2, J_1)$ or $\mathfrak{S}(\Omega_4, J_1)$. Any restricted Lie algebra of order 3 in the classes enumerated here is isomorphic to a $\mathfrak{B}'_l$. Our determination of the automorphism groups gives the well-known isomorphism theorems between certain projective orthogonal and full projective groups.[23] Aside from the cases we have noted here, no other isomorphisms exist between the Lie algebras enumerated here. For any isomorphism may be extended to an isomorphism between algebras $\mathfrak{S}(\Omega_n, J_i)$ or between these and Ω'_{ml}'s. Whether or not this holds for the groups of automorphisms is an open question.

---

[23] Cf. van der Waerden, *Gruppen von linearen Transformationen*, Berlin, 1935, pp. 18-28.

3.2. Extension to non-normal algebras. If $\mathfrak{L}$ is an (ordinary) simple Lie algebra over Φ the enveloping algebra of the linear transformations $x \to [x, a] = xA$ is a matrix algebra Σ_m where Σ is a field of finite order over Φ.[24] $\mathfrak{L}$ may be regarded as an algebra over Σ and when this is done ($\mathfrak{L}$ over $\Sigma)_\Omega$, Ω the algebraic closure of Σ, is simple. Σ is the only field of operators containing Φ for which this holds. It is known that any automorphisms S of $\mathfrak{L}$ over Φ induces an automorphism s in Σ over Φ such that $(x\xi)^S = x^S \xi^s$ if $\xi \epsilon \Sigma$. Now consider a derivation D in $\mathfrak{L}$ over Φ. If the A_i are defined as above: $[x, a_i] = xA_i$ then

$$x(\Sigma A_1A_2 \cdots A_r)D = (xD) \Sigma A_1A_2 \cdots A_r + x(\Sigma A'_1A_2 \cdots A_r + \Sigma A_1A'_2 \cdots A_r + \cdots + \Sigma A_1A_2 \cdots A'_r)$$

where A'_i is the transformation $x \to [x, a_iD]$. Hence if $\Sigma A_1A_2 \cdots A_r = \Sigma B_1B_2 \cdots B_s$ then

$$\Sigma A'_1A_2 \cdots A_r + \cdots + \Sigma A_1A_2 \cdots A'_r = \Sigma B'_1B_2 \cdots B_s + \cdots + \Sigma B_1B_2 \cdots B'_s.$$

Thus the correspondence

$$\Sigma A_1A_2 \cdots A_1 \to \Sigma A'_1A_2 \cdots A_r + \cdots + \Sigma A_1A_2 \cdots A'_r$$

is a derivation in the associated algebra Σ_m and induces a derivation d in Σ over Φ such that $(x\xi)D = (xD)\xi + x(\xi^d)$. It follows that the derivation algebra $\mathfrak{D}$ of $\mathfrak{L}$ over Φ contains as an ideal $\mathfrak{J}$ the set of derivations of $\mathfrak{L}$ over Σ and that $\mathfrak{D}/\mathfrak{J}$ is isomorphic to a subalgebra of the algebra of derivations of the field Σ over Φ.

For example let $\Sigma = \Phi(b_1, b_2, \cdots, b_m)$ where $b_i{}^p = \beta_i$ is in Φ and $\mathfrak{L} = \Sigma'_{nl}$ with $n \not\equiv 0 \pmod{p}$. If d is any derivation in Σ over Φ the correspondence $(\xi_{ij}) \to (\xi_{ij}{}^d)$ is a derivation in $\mathfrak{L}$ over Φ which induces the derivation d in Σ. Hence $\mathfrak{D}/\mathfrak{J}$ is isomorphic to the complete algebra of derivations of Σ over Φ. It is known that the latter is a simple Lie algebra.[25] Since we have shown above that $\mathfrak{J}$ is the algebra of inner derivations of Σ'_{nl}, the outer derivation algebra of $\mathfrak{L}$ is not solvable and this contradicts a conjecture recently made by Zassenhaus.[2]

UNIVERSITY OF NORTH CAROLINA
ON LEAVE AT
THE JOHNS HOPKINS UNIVERSITY.

---

[24] See the author's "A note on non-associative algebras," *Duke Mathematical Journal*, vol. 3 (1937), pp. 544-548.

[25] See [6], p. 218.

ON THE SECOND VARIATION IN CERTAIN ANORMAL PROBLEMS OF THE CALCULUS OF VARIATIONS.*

By E. J. McShane.

In a previous paper in this *Journal*[1] I showed that if a curve $y = y(x)$ gives to an integral $\int f(x, y, y')\,dx$ a strong relative minimum in the class of curves satisfying certain differential equations $\phi_\alpha(x, y, y') = 0$ and certain end conditions, there exist multipliers $\lambda_0 \geqq 0, \lambda_\alpha(x)$ such that for the combination $\lambda_0 f + \lambda_\alpha \phi_\alpha$ the Du Bois–Reymond equations, the transversality condition, the Weierstrass condition and the Clebsch condition are all satisfied. In the present note we shall investigate the possibility of choosing the multipliers in such a way that the fourth necessary condition (the Jacobi, or Mayer, condition) is also satisfied, by which we mean that the second variation is non-negative. The results here attained, unlike those of the previous paper, are not free of normality assumptions. It is shown that if the order of anormality is 0 or 1, the choice is possible; but if the order of anormality exceeds 1, there may be no set of multipliers with which the second variation is non-negative. In problems involving a single differential equation $\phi_\alpha = 0$ the order of anormality cannot exceed 1; so for such problems we may reasonably consider that the theory is now as completely developed as the theory of free problems.

For brevity, we shall retain the notation and hypotheses of MLP, and shall frequently refer to its contents.

1. Weak and strong variations. Let us recall that the Lagrange problem which we are studying is that of minimizing an integral

$$(1) \qquad I = \int_{x_1}^{x_2} f(x, y, y')\,dx$$

in the class of D' curves

$$(2) \qquad y_i = y_i(x) \qquad (x_1 \leqq x \leqq x_2;\ i = 1, \cdots, n)$$

which satisfy a system of differential equations

$$(3) \qquad \phi_\alpha(x, y, y') = 0 \qquad (\alpha = 1, \cdots, m < n)$$

* Received March 19, 1940; Revised February 27, 1941.

[1] "On multipliers for Lagrange problems," *American Journal of Mathematics*, vol. 61 (1939), pp. 809-819; henceforth cited as MLP.

and end conditions

$$\psi_\mu(x_1, y(x_1), x_2, y(x_2)) = 0 \qquad (\mu = 1, \cdots, p \leqq 2n+2). \tag{4}$$

The hypotheses concerning the continuity and independence of the functions ϕ_α, ψ_μ and f are the usual ones.[2] We suppose that a curve

$$E_{12} : y_i = y_i(x), \qquad (x_1 \leqq x \leqq x_2;\ i = 1, \cdots, n) \tag{5}$$

gives a strong relative minimum to I in the class of curves satisfying (3) and (4), and seek the conditions which it must satisfy.

By a *weak variation* we mean a system $[\xi_1, \xi_2, \eta(x)]$, where ξ_1 and ξ_2 are arbitrary numbers and $\eta_1(x), \cdots, \eta_n(x)$ are functions of class D' on the interval $x_1 \leqq x \leqq x_2$ which satisfy the equations of variation of (3). By a *strong variation* we mean a system $[X, Y']$ such that $(X, y(X))$ is a point of E_{12} which is not an end or corner of E_{12} and $(X, y(X), Y')$ is admissible. In MLP, § 1, we saw that given any finite set of weak variations

$$[\xi_{1,j}, \xi_{2,j}, \eta_j(x)], \qquad (j = 1, \cdots, w)$$

and strong variations

$$[X_k, Y'_k] \qquad (k = 1, \cdots, s)$$

it is possible to construct a family of curves

$$y_i = y_i(x, b_1, \cdots, b_w, e_1, \cdots, e_s), \qquad x_1(b, e) \leqq x \leqq x_2(b, e) \tag{6}$$

defined for $|b_j| \leqq \epsilon$ and $0 \leqq e_k \leqq \epsilon$ (ϵ a small positive number), satisfying the differential equations (3) and containing E_{12} for $b = e = 0$. We suppose that the reader is familiar with this construction. To each curve (6) there corresponds a set of values of the functions

$$\begin{aligned} \rho_0(b, e) &= \int_{x_1(b,e)}^{x_2(b,e)} f(x, y(x, b, e), y'(x, b, e))\,dx - I(E_{12}), \\ \rho_\mu(b, e) &= \psi_\mu(x_1(b, e), y(x_1(b, e), b, e), x_2(b, e), y(x_2(b, e), b, e)) \\ &\qquad (\mu = 1, \cdots, p). \end{aligned} \tag{7}$$

The derivatives of these functions have been computed in MLP; for the moment, we do not need to know what these derivatives are.

In this paper we shall be interested in the second variation of the integral. There is no difficulty in computing the second-order partial derivatives of the $\rho_i(b, e)$ with respect to the b_j if we assume that

(8) *for each value of $s (s = 1, 2)$, either the end conditions* (4) *fix x_s, or else $y(x)$ is of class C'' near x_s.*

---

[2] See, for instance, G. A. Bliss, "The problem of Lagrange in the calculus of variations," *American Journal of Mathematics*, vol. 52 (1930), pp. 673-744.

As is usual in the calculus of variations, we restrict our attention to the second variations due to weak variations $[\xi_1, \xi_2, \eta(x)]$ which satisfy the conditions

(9) $$\Psi_\mu(\xi_1, \eta(x_1), \xi_2, \eta(x_2)) = 0 \qquad (\mu = 1, \cdots, p).$$

Here the equations $\Psi_\mu = 0$ are the equations of variation [3] of (4). In terms of the functions ρ_i, if a weak variation $[\xi_{1,j}, \xi_{2,j}, \eta_j(x)]$ satisfies (9) and b_j is the corresponding parameter, the equations

(10) $$\partial\rho_\mu/\partial b_j = 0 \qquad (\mu = 1, \cdots, p)$$

hold at $b = e = 0$. We define the aggregate W of *second variation vectors* as follows.

(11) *W is the aggregate of all vectors $(\partial^2\rho_0/\partial b_j^2, \cdots, \partial^2\rho_p/\partial b_j^2)$ evaluated at $b = e = 0$ corresponding to weak variations $[\xi_{1,j}, \xi_{2,j}, \eta_j(x)]$ which satisfy equations* (9) (and hence satisfy (10)).

2. Combinations of second variations. Let $w_j = (w_{0,j}, \cdots, w_{p,j})$, $(j = 1, 2)$, be two vectors of the class W, corresponding, respectively, to the weak variations $[\xi_{1,j}, \xi_{2,j}, \eta_j(x)]$. Corresponding to these variations we construct a family

$$y_i = y_i(x, b_1, b_2), \qquad x_1(b_1, b_2) \leqq x \leqq x_2(b_1, b_2)$$

of curves satisfying the differential equations (3) and containing E_{12} for $b_1 = b_2 = 0$. These curves define functions $\rho_i(b_1, b_2)$ as in (7). For an arbitrary θ we now compute the vector whose components are

(12) $$w_i(\theta) = \partial^2\rho_i(r\cos\theta, r\sin\theta)/\partial r^2,$$

evaluated at $r = 0$. We find at once that

(13) $$w_i(\theta) = w_{i,1}\cos^2\theta + 2a_i\cos\theta\sin\theta + w_{i,2}\sin^2\theta,$$

where

(14) $$a_i = \partial^2\rho_i/\partial b_1\partial b_2\,|_{b=0}.$$

Equation (13) can be written in the form

(15) $$w_i(\theta) = 1/2(w_{i,1} + w_{i,2}) + 1/2(w_{i,1} - w_{i,2})\cos 2\theta + a_i\sin 2\theta.$$

Clearly

(16) $$w_i(0) = w_i(\pi) = w_{i,1}, \qquad w_i(\pi/2) = w_{i,2}.$$

We are particularly interested in the motion of the point $(w_0(\theta), w_1(\theta))$

---

[3] *Loc. cit.*[2], equation (44).

for $0 \leqq \theta \leqq \pi$. This point moves continuously as a function of θ. Suppose first that the determinant

$$(17) \qquad \begin{vmatrix} 1/2(w_{0,1} - w_{0,2}) & a_0 \\ 1/2(w_{1,1} - w_{1,2}) & a_1 \end{vmatrix}$$

is not zero. Then the first two of equations (15) can be solved for $\cos 2\theta$ and $\sin 2\theta$; these are linear functions of w_0 and w_1. If we square and add, we find that w_0, w_1 satisfy an equation of the form

$$\textit{positive definite quadratic function of } (w_0, w_1) = 1.$$

Thus $(w_0(\theta), w_1(\theta))$ describes an ellipse. Had we first translated axes, bringing the origin to the point

$$P: (1/2(w_{0,1} + w_{0,2}), \qquad 1/2(w_{1,1} + w_{1,2})),$$

the quadratic function would have been homogeneous. Hence P is the center of the ellipse. Therefore as θ transverses the interval $0 \leqq \theta \leqq \pi$ the point $(w_0(\theta), w_1(\theta))$ traverses an ellipse of which the segment joining $(w_{0,1}, w_{1,1})$ and $(w_{0,2}, w_{1,2})$ is a diameter.

If the determinant (17) vanishes, a linear combination of

$$w_0(\theta) - 1/2(w_{0,1} + w_{0,2}) \text{ and } w_1(\theta) - 1/2(w_{1,1} + w_{1,2})$$

vanishes identically. In this case the point $(w_0(\theta), w_1(\theta))$ moves in such a way as to remain on the line through $(w_{0,1}, w_{1,1})$ and $(w_{0,2}, w_{1,2})$. It is convenient to regard the locus of $(w_0(\theta), w_1(\theta))$ as a degenerate ellipse, twice covering a segment of that line.

Since for $r = 0$ we have

$$(18) \quad \partial\rho_\mu/\partial r(r\cos\theta, r\sin\theta) = (\partial\rho_\mu/\partial e_1)\cos\theta + (\partial\rho_\mu/\partial e_2)\sin\theta = 0, \qquad (\mu = 1, \cdots, p)$$

the vector $w(\theta)$ belongs to the class W.

3. A convex set defined by the variations. Henceforth we make the assumption that

(19) *The order of anormality of E_{12} does not exceed* 1.

That is we assume the existence of $p - 1$ weak variations[4]

$$(20) \qquad [\xi_{1,l}, \xi_{2,l}, \eta_l(x)], \qquad (l = 2, 3, \cdots, p)$$

---

[4] The principal theorem of this note would still be valid if we replaced hypothesis (19) by the weaker assumption that there are variations, weak or strong, such that the linear combinations of the corresponding vectors ρ_i with non-negative coefficients cover a subspace $u_0 = a_2u_2 + \cdots + a_pu_p$, $u_1 = b_2u_2 + \cdots + b_pu_p$. But this weakening of hypothesis would entail a complication of the proof.

such that the resulting vectors

$$\rho_{j,l} = [\partial \rho_j / \partial b_l]_{b=0}$$

cause the matrix

$$\| \rho_{\mu,l} \| \qquad (\mu = 1, \cdots, p;\ l = 2, \cdots, p)$$

to have rank $p - 1$. There is no loss of generality in assuming that the determinant

$$(21) \qquad | \rho_{h,l} | \qquad (h, l = 2, \cdots, p)$$

is different from zero.

Now let the set K_+ consist of all vectors ρ_j defined by weak or strong variations, all vectors w_j of the class W, and all linear combinations

$$b_1 \rho_{j,1} + \cdots + b_k \rho_{j,k} + c_1 w_{j,1} + \cdots + c_h w_{j,h}$$

of such vectors with non-negative coefficients b_i and c_i. It is easy to see that K_+ is convex, so that its closure $\bar{K}_+$ is also convex. Also, K_+ is a cone with vertex at the origin; that is, if ρ_j belongs to K_+, so does $b\rho_j$ for all non-negative b. Hence $\bar{K}_+$ is also a cone with vertex at the origin. Concerning K_+ we shall prove the following lemma.

LEMMA. *If the hypotheses* (8) *and* (19) *are satisfied, the point* $-\delta_0 \equiv (-1, 0, \cdots, 0)$ *of* $(p+1)$*-space is not interior to* $\bar{K}_+$.

This lemma is the essential part of this note; its proof constitutes §§ 4-7.

4. An embedding family. Suppose that the conclusion of the lemma is false. Then a neighborhood of $-\delta_0$ is interior to $\bar{K}_+$. We choose a simplex with vertices $P_1, \cdots, P_{p+2}$ lying in this neighborhood and containing $-\delta_0$ in its interior. Arbitrarily near P_j $(j = 1, \cdots, p+2)$ there are points P'_j of K_+. Hence these points P'_j can be so chosen that they are the vertices of a simplex containing $-\delta_0$ in its interior. Since the points P'_j belong to the convex set K_+, the entire simplex $P'_1 \cdots P'_{p+2}$ belongs to K_+. Hence $-\delta_0$ is interior to K_+. It follows that if ϵ is a sufficiently small positive number, the points $(-1, \epsilon, 0, \cdots, 0)$ and $(-1, -\epsilon, 0, \cdots, 0)$ both belong to K_+.

By definition of K_+, there are vectors $w_{i,1}, \cdots, w_{i,q}$ of W and weak or strong variation vectors $\rho_{i,p+1}, \cdots, \rho_{i,p+r}$ such that

$$(22) \qquad \begin{aligned} -1 &= A_1 w_{0,1} + \cdots + A_q w_{0,q} + B_{p+1} \rho_{0,p+1} + \cdots + B_{p+r} \rho_{0,p+r}, \\ \epsilon &= A_1 w_{1,1} + \cdots + A_q w_{1,q} + B_{p+1} \rho_{1,p+1} + \cdots + B_{p+r} \rho_{1,p+r}, \\ 0 &= A_1 w_{l,1} + \cdots + A_q w_{l,q} + B_{p+1} \rho_{l,p+1} + \cdots + B_{p+r} \rho_{l,p+r}, \\ &\qquad (l = 2, \cdots, p), \end{aligned}$$

the A_i and B_j being non-negative. Likewise there are vectors $w^*_{i,1}, \cdots, w^*_{i,s}$ of W and weak or strong variation vectors $\rho^*_{i,1}, \cdots, \rho^*_{i,t}$ such that

$$
\begin{aligned}
-1 &= C_1 w^*_{0,1} + \cdots + C_s w^*_{0,s} + D_1 \rho^*_{0,1} + \cdots + D_t \rho^*_{0,t}, \\
(23) \qquad \epsilon &= C_1 w^*_{1,1} + \cdots + C_s w^*_{1,1} + D_1 \rho^*_{1,1} + \cdots + D_t \rho^*_{1,t}, \\
0 &= C_1 w^*_{l,1} + \cdots + C_s w^*_{l,s} + D_1 \rho^*_{l,1} + \cdots + D_t \rho^*_{l,t}, \\
& \qquad\qquad\qquad\qquad\qquad\qquad (l = 2, \cdots, p),
\end{aligned}
$$

the C_i and D_j being non-negative. We now construct a family of curves

$$
(24) \qquad y_i = y_i(x, e_1, \cdots, e_q, e^*_1, \cdots, e^*_s, b_2, \cdots, b_{p+r}, b^*_1, \cdots, b^*_t), \\
x_1(e, e^*, b, b^*) \leqq x \leqq x_2(e, e^*, b, b^*),
$$

incorporating the variations involved in (22) and (23) and also the variations (20). The functions in (24) are defined on a set

$$
\begin{aligned}
(25) \qquad & |e_i| \leqq \delta, \ |e^*_i| \leqq \delta, \ |b| \leqq \delta & (l = 2, \cdots, p), \\
& 0 \leqq b_h \leqq \delta & (h = p+1, \cdots, p+r), \\
& 0 \leqq b^*_k \leqq \delta & (k = 1, \cdots, t),
\end{aligned}
$$

where δ is positive. (The first three sets of parameters correspond to weak variations, and hence we need not restrict their sign.) The corresponding functions $\rho_i(e, e^*, b, b^*)$, defined as in (7), have derivatives at $e = e^* = b = b^* = 0$ which satisfy the following equations.

$$
(26) \qquad \partial\rho_\mu/\partial e_j = \partial\rho_\mu/\partial e^*_k = 0 \qquad (\mu = 1, \cdots, p).
$$

(For the vectors w_j, w^*_k belong to W, so that equations (10) hold).

$$
\begin{aligned}
(27) \qquad & \partial^2\rho_i/\partial e_j^2 = w_{i,j} & (i = 0, \cdots, p;\ j = 1, \cdots, q) \\
(28) \qquad & \partial^2\rho_i/\partial e^{*2}_k = w^*_{i,k} & (i = 0, \cdots, p;\ k = 1, \cdots, s) \\
(29) \qquad & \partial\rho_i/\partial b_j = \rho_{i,j} & (i = 0, \cdots, p;\ j = 2, \cdots, p+r) \\
(30) \qquad & \partial\rho_i/\partial b^*_k = \rho^*_{i,k} & (i = 0, \cdots, p;\ k = 1, \cdots, t).
\end{aligned}
$$

We may assume the number δ in (25) to be small enough so that the curves (24) lie in the neighborhood on which E_{12} minimizes the integral I. Then the conditions

$$
(31) \qquad \rho_0(e, e^*, b, b^*) < 0, \qquad \rho_\mu(e, e^*, b, b^*) = 0 \qquad (\mu = 1, \cdots, p)
$$

are incompatible. For the curves (24) lie in the neighborhood on which E_{12} minimizes I. They satisfy the differential equations (3), and if the last p equations (31) hold they satisfy the end-conditions (4), because of definition (7). Hence the integral along the curve (24) can not be less than $I(E_{12})$; that is, by (7), $\rho_0(e, e^*, b, b^*)$ can not be negative.

5. Change of coördinates. The vectors $(\rho_0, \cdots, \rho_p)$ lie in a $(p+1)$-dimensional space of points $(u_1, \cdots, u_p)$. The vectors $(1, 0, \cdots, 0)$,

$(0,1,0,\cdots,0)$ and $(\rho_{0,l},\cdots,\rho_{p,l})$ $(l=2,\cdots,p)$ are linearly independent, since the determinant (21) is not zero. Hence we can use these $p+1$ vectors as a basis for the space. That is, we introduce an affine transformation

$$z_i = M_{i,j}u_j \qquad (i,j=0,1,\cdots,p)$$

which leaves $\delta_{i,0}$ and $\delta_{i,1}$ unchanged and maps $u_i=\rho_{i,l}$ on $\delta_{i,l}$ $(l=2,\cdots,p)$. Here $\delta_{i,j}$ is the Kronecker δ, with value 1 if the subscripts are equal and 0 otherwise. The matrix $M_{i,j}$ is clearly non-singular. All points in the u_0u_1-plane are invariant under the mapping. The functions

$$z_i(e,e^*,b,b^*) = M_{i,j}\rho_j(e,e^*,b,b^*)$$

are readily seen to satisfy the following conditions.

$$\partial z_\mu/\partial e_j = \partial z_\mu/\partial e^*_k = 0 \qquad (\mu=1,\cdots,p). \tag{32}$$

(For by (26) the vectors $\partial\rho_i/\partial e_j$ and $\partial\rho_i/\partial e^*_k$ are in the u_0u_1-plane, and are, therefore, invariant).

$$\partial^2 z_i/\partial e_j{}^2 = v_{i,j} \equiv M_{i,h}w_{h,j} \qquad (i=0,\cdots,p;\ j=1,\cdots,q) \tag{33}$$

$$\partial^2 z_i/\partial e^*_k{}^2 = v^*_{i,k} \equiv M_{i,h}w^*_{h,k} \qquad (i=0,\cdots,p;\ k=1,\cdots,t) \tag{34}$$

$$\partial z_i/\partial b_j = z_{i,j} \equiv M_{i,h}\rho_{h,j} \qquad (i=0,\cdots,p;\ j=2,\cdots,p+r) \tag{35}$$

$$\partial z_i/\partial b^*_k = z^*_{i,k} \equiv M_{i,h}\rho^*_{h,k} \qquad (i=0,\cdots,p;\ k=1,\cdots,t). \tag{36}$$

Recalling that $(-1,\epsilon,0,\cdots,0)$ is mapped on itself, we obtain from (22)

$$-1 = A_1v_{0,1}+\cdots+A_qv_{0,q}+B_{p+1}z_{0,p+1}+\cdots+B_{p+r}z_{0,p+r}, \tag{37}$$

$$\epsilon = A_1v_{1,1}+\cdots+A_qv_{1,q}+B_{p+1}z_{1,p+1}+\cdots+B_{p+r}z_{0,p+r}. \tag{38}$$

Likewise, by (23),

$$-1 = C_1v^*_{0,1}+\cdots+C_sv^*_{0,s}+D_1z^*_{0,1}+\cdots+D_tz^*_{0,t}, \tag{39}$$

$$-\epsilon = C_1v^*_{1,1}+\cdots+C_sv^*_{1,s}+D_1z^*_{0,1}+\cdots+D_tz^*_{0,t}. \tag{40}$$

By the choice of axes, with (35), we have

$$z_{i,l} = \delta_{il} \qquad (i=0,\cdots,p;\ l=2,\cdots,p). \tag{41}$$

By the invariance of points in the (u_0,u_1)-plane, the incompatibility of conditions (31) implies

The conditions

$$z_0(e,e^*,b,b^*) < 0 = z_\mu(e,e^*,b,b^*) \qquad (\mu=1,\cdots,p) \tag{42}$$

are incompatible.

6. Reduction in the number of parameters. The number of parameters (e, e^*, b, b^*) is inconveniently large, so we shall show that it can be reduced without harm.

Suppose that more than one of the coefficients A_i in (37) is different from zero; to be specific, suppose A_1 and A_2 are both positive. We set $e_1 = r\cos\theta$, $e_2 = r\sin\theta$ and set all other parameters equal to 0. The point in the (z_0, z_1)-plane with coördinates

$$v_j(\theta) = \frac{\partial^2}{\partial r^2} z_j(r\cos\theta,\ r\sin\theta,\ 0,\cdots,0)\,|_{r=0}, \qquad (j = 0, 1)$$

traverses an ellipse as θ varies from 0 to π, as we saw in § 2. The segment joining $(v_{0,1}, v_{1,1})$ and $(v_{0,2}, v_{1,2})$ is a diameter of this ellipse. Some point of this segment lies on the ray from the origin through the point Q: $(A_1v_{0,1} + A_2v_{0,2},\ A_1v_{1,1} + A_2v_{1,2})$. This ray, considered as starting at the origin, departs from the ellipse at a point $(v_0(\bar\theta), v_1(\bar\theta))$. Hence this point is collinear with Q and the origin, and there is a positive constant A_0 such that

$$\begin{aligned} A_0v_0(\bar\theta) &= A_1v_{0,1} + A_2v_{0,2}, \\ A_0v_1(\bar\theta) &= A_1v_{1,1} + A_2v_{1,2}. \end{aligned} \tag{43}$$

Since $\bar\theta$ is constant, we have replaced the two parameters e_1, e_2 by a single parameter r, diminishing by one the number of vectors $w_{i,j}$ in equations (37) and (38).

Continuing this process, we finally find that equations (37) and (38) can be satisfied with $q = 1$, that is, with the use of only one vector $v_i = M_{ij}w_j$ having w in the class W. We drop the corresponding subscript; this single vector we call v_i instead of $v_{i,1}$, and its corresponding parameter we call e.

In a like manner equations (39) and (40) can be satisfied with the use of the transform v^*_j of a single vector w^*_j of W. The corresponding parameter we denote by e^*. Now equations (32), (33) and (34) hold with subscripts j, k omitted; equations (35) and (36) hold without change.

In equations (37) and (38) there is now a single vector v, with coefficient $A \geqq 0$. Likewise in (39) and (40) there is a single vector v^* with coefficient $C \geqq 0$. To all the parameters except $b_2, \cdots, b_p$ in (24) we assign values in terms of two parameters σ, τ as follows.

$$\begin{aligned} e &= A^{\frac12}\sigma, \quad b_h = 1/2\, B_h\sigma^2 && (h = p+1,\cdots,p+r), \\ e^* &= C^{\frac12}\tau, \quad b^*_k = 1/2\, D_k\tau^2 && (k = 1,\cdots,t). \end{aligned} \tag{44}$$

Then at $\sigma = \tau = b = 0$ we have by (32), (33), (35), (37) and (38)

$$\partial z_\mu/\partial\sigma = \partial z_\mu/\partial\tau = 0 \qquad (\mu = 1,\cdots,p). \tag{45}$$

$$(46)\qquad \partial^2 z_j/\partial\sigma^2 = A\partial^2 z_j/\partial e^2 + \sum_{h=p+1}^{p+r} B_h \partial z_j/\partial b_h$$
$$= \begin{cases} -1 & \text{if } j=0 \\ \epsilon & \text{if } j=1, \end{cases}$$

and likewise

$$(47)\qquad \partial^2 z_j/\partial\tau^2 = \begin{cases} -1 & \text{if } j=0, \\ -\epsilon & \text{if } j=1. \end{cases}$$

If τ, σ are near enough to 0, the parameters defined in (44) satisfy (25).

Summing up the results so far, we now have a $(p+1)$-parameter family

$$(48)\qquad y_i = \bar{y}_i(x, \sigma, \tau, b_2, \cdots, b_p), \qquad \bar{x}_1(\sigma, \tau, b) \leqq x \leqq \bar{x}_2(\sigma, \tau, b)$$

defined for

$$(49)\qquad |\sigma| \leqq \delta_1, \quad |\tau| \leqq \delta_1, \quad |b| \leqq \delta \qquad (l = 2, \cdots, p)$$

(where δ_1 is a sufficiently small positive number) such that the corresponding functions $z_i(\sigma, \tau, b)$ satisfy the following conditions at $\sigma = \tau = b = 0$.

$$(50)\qquad \partial z_\mu/\partial\sigma = \partial z_\mu/\partial\tau = 0 \qquad (\mu = 1, \cdots, p),$$

$$(51)\qquad \partial^2 z_0/\partial\sigma^2 = \partial^2 z_0/\partial\tau^2 = -1,$$

$$(52)\qquad \partial^2 z_1/\partial\sigma^2 = -\partial^2 z_1/\partial\tau^2 = \epsilon,$$

$$(53)\qquad \partial z_i/\partial b_l = \delta_{il} \qquad (i = 0, \cdots, p;\ l = 2, \cdots, p).$$

(The last statement follows from (35) and (41)). Moreover, by (31) the conditions

$$(54)\qquad z_0(\sigma, \tau, b) < 0, \qquad z_\mu(\sigma, \tau, b) = 0 \qquad (\mu = 1, \cdots, p)$$

are incompatible.

The equations

$$(55)\qquad z_l(\sigma, \tau, b_2, \cdots, b_p) = 0 \qquad (l = 2, \cdots, p)$$

have the initial solution $\sigma = \tau = b = 0$. At this solution the jacobian with respect to the b_h has value 1, by (41). Hence equations (55) have solutions

$$(56)\qquad b_l = b_l(\sigma, \tau), \qquad (l = 2, \cdots, p)$$

defined and of class C'' near $(0, 0)$. For compactness we define

$$(57)\qquad U_i(\sigma, \tau) = z_i(\sigma, \tau, b_2(\sigma, \tau), \cdots, b_p(\sigma, \tau)), \qquad (i = 0, \cdots, p).$$

Then

$$(58)\qquad U_l(\sigma, \tau) = 0 \qquad (l = 2, \cdots, p)$$

for all (σ, τ) near $(0, 0)$.

In (58) we differentiate with respect to σ and set $\sigma = \tau = 0$. We find, with (41),

(59) $$\partial z_l/\partial\sigma + \delta_{lj}(\partial b_j/\partial\sigma) = 0.$$

By (45), this implies

(60) $$\partial b_l/\partial\sigma = 0 \qquad (l = 2, \cdots, p;\ \sigma = \tau = 0).$$

In the same way

(61) $$\partial b_l/\partial\tau = 0 \qquad (l = 2, \cdots, p;\ \sigma = \tau = 0).$$

If we differentiate U_j $(j = 0, 1)$ twice with respect to σ and set $\sigma = \tau = 0$ we obtain, by (41) and (60),

(62) $$\partial^2 U_j/\partial\sigma^2 = \partial^2 z_j/\partial\sigma^2, \qquad (j = 0, 1).$$

A corresponding equation holds for the second derivative with respect to τ. Hence at $\sigma = \tau = 0$

(63) $$\partial^2 U_0/\partial\sigma^2 = \quad \partial^2 U_0/\partial\tau^2 = -1,$$

(64) $$\partial^2 U_1/\partial\sigma^2 = -\partial^2 U_1/\partial\tau^2 = \epsilon.$$

Also by (50) and (60),

(65) $$\partial U_1/\partial\sigma = \partial U_1/\partial\tau = 0 \qquad (\sigma = \tau = 0).$$

Finally

(66) *the conditions* $U_0(\sigma, \tau) < 0$, $U_1(\sigma, \tau) = 0$ *are incompatible.*

For if (66) holds, by (58) the conditions (54) are all satisfied, and these are incompatible.

7. Completion of proof of lemma. The parameters σ, τ are not restricted as to sign, so we may replace either one by its negative without leaving the range of definition or affecting any one of the statements (62) to (66). Therefore we may assume that at $\sigma = \tau = 0$

(67) $$\partial U_0/\partial\sigma \leqq 0,$$

(68) $$\partial U_0/\partial\tau \leqq 0.$$

Let us define

(69) $$u_j(\theta) = \partial^2 U_j(r\cos\theta,\ r\sin\theta)/\partial r^2 \qquad (j = 0, 1),$$

the derivatives being evaluated at $r = 0$. As we saw in § 2, when θ traverses the interval $0 \leqq \theta \leqq \pi/2$ the point $(u_0(\theta), u_1(\theta))$ describes a semi-ellipse from $(-1, \epsilon)$ to $(-1, -\epsilon)$; when θ traverses the interval $\pi/2 \leqq \theta \leqq \pi$ the point $(u_0(\theta), u_1(\theta))$ describes the other half of the same ellipse from $(-1, -\epsilon)$ to $(-1, \epsilon)$. Thus there are two values of θ, one in the interval $(0, \pi/2)$ and the other in the interval $(\pi/2, \pi)$, for which $u_1(\theta)$ vanishes.

We distinguish two cases.

CASE I. *The sign* $<$ *holds in one of the inequalities* (67), (68).

In this case we denote by $\bar{\theta}$ the value of θ between 0 and $\pi/2$ for which

$$u_1(\bar{\theta}) = 0. \tag{70}$$

It follows at once that at $r = 0$

$$\frac{\partial}{\partial r} U_0(r\cos\bar{\theta},\ r\sin\bar{\theta}) = (\partial U_0/\partial\sigma)\cos\bar{\theta} + (\partial U_0/\partial\tau)\sin\bar{\theta} < 0. \tag{71}$$

CASE II. *The sign = holds both in* (67) *and in* (68). The computation leading to (71) yields, at $r = 0$,

$$\frac{\partial}{\partial r} U_0(r\cos\theta,\ r\sin\theta) = 0, \qquad (0 \leqq \theta \leqq \pi). \tag{72}$$

Since the center of the ellipse $u_0 = u_0(\theta)$, $u_1 = u_1(\theta)$ is the point $(-1, 0)$, at least one of the intersections of this ellipse with the u_0-axis must lie on the negative half of that axis. We choose $\bar{\theta}$ to be such that

$$u_1(\bar{\theta}) = 0, \qquad u_0(\bar{\theta}) < 0. \tag{73}$$

By continuity, there exists a positive number γ such that, in Case I,

$$\frac{\partial}{\partial r} U_0(r\cos\theta,\ r\sin\theta) < 0 \qquad (0 \leqq r \leqq \gamma,\ |\theta - \bar{\theta}| \leqq \gamma), \tag{74}$$

and in Case II (recalling (69) and (73))

$$\frac{\partial^2}{\partial r^2} U_0(r\cos\theta,\ r\sin\theta) < 0 \qquad (0 \leqq r \leqq \gamma,\ |\theta - \bar{\theta}| \leqq \gamma). \tag{75}$$

Equation (15), with $i = 1$ and u in place of w, yields

$$u_1(\theta) = \epsilon\cos 2\theta + a_1\sin 2\theta. \tag{76}$$

Hence

$$u'_1(\theta) = -2\epsilon\sin 2\theta + 2a_1\cos 2\theta. \tag{77}$$

From these equations,

$$4[u_1(\theta)]^2 + [u'_1(\theta)]^2 = 4\epsilon^2 + 4a_1^2 > 0. \tag{78}$$

This, with (70) or (73), implies

$$u'_1(\bar{\theta}) \neq 0. \tag{79}$$

Therefore $u_1(\theta)$ changes sign on passing through $\bar{\theta}$, and, by diminishing γ if necessary, we can bring it about that $u_1(\bar{\theta} + \gamma)$ and $u_1(\bar{\theta} - \gamma)$ have opposite signs. To be specific we suppose

$$u_1(\bar{\theta} - \gamma) < 0 < u_1(\bar{\theta} + \gamma). \tag{80}$$

By (65), we have at $r=0$

$$\partial U_1(r\cos\theta,\ r\sin\theta)/\partial r = 0 \qquad (0 \leqq \theta \leqq \pi). \tag{81}$$

Recalling (69) and (80), we find by Taylor's theorem that for all r near 0 the inequalities

$$\begin{aligned} U_1(r\cos(\bar\theta-\gamma),\ r\sin(\bar\theta-\gamma)) & \\ < 0 < U_1(r\cos(\bar\theta+\gamma),\ & r\sin(\bar\theta+\gamma)) \end{aligned} \tag{82}$$

hold. Since U_1 is continuous, this implies

(83) *For each sufficiently small positive number r there is a $\theta(r)$ such that $|\theta(r)-\bar\theta|<\gamma$ and*

$$U_1(r\cos\theta(r),\ r\sin\theta(r)) = 0.$$

For such r we now compute U_0. We may suppose $0<r<\gamma$. Then in Case I we have, by the mean value theorem,

$$U_0(r\cos\theta(r),\ r\sin\theta(r)) = r\frac{\partial}{\partial\rho}U_0(\rho\cos\theta(r),\rho\sin\theta(r)),$$

the derivative on the right being computed for some number ρ such that $0<\rho<r$. This, with the inequality (71), implies

$$U_0(r\cos\theta(r),\ r\sin\theta(r)) < 0. \tag{84}$$

In Case II we compute U_0 by Taylor's formula, recalling (72). We find

$$U_0(r\cos\theta(r),\ r\sin\theta(r)) = \tfrac{1}{2}r^2\frac{\partial^2}{\partial\rho^2}U_0(\rho\cos\theta(r),\ \rho\sin\theta(r)). \tag{85}$$

By (75), in this case too inequality (84) holds. But, by (66), the statements (83) and (84) are incompatible. Hence we have at last reached the desired contradiction, and the lemma is established.

8. The Jacobi condition. Our lemma being established, we can as in MLP (p. 815) choose numbers $\lambda_0 \geqq 0$, $d_1,\cdots,d_p$, not all zero, such that for every point $(u_0,\cdots,u_p)$ of K_+ the inequality

$$\lambda_0 u_0 + d_\mu u_\mu \geqq 0 \tag{86}$$

is satisfied. From this all the statements of § 3 of MLP follow; in particular, the existence of multipliers $\lambda_0, \lambda_a(x)$ such that for

$$F(x,y,y',\lambda) \equiv \lambda_0 f(x,y,y') + \lambda_a\phi_a(x,y,y')$$

the Du Bois–Reymond relation, the transversality condition, the Weierstrass condition and the Clebsch condition all hold.

Suppose now that $[\xi_1, \xi_2, \eta_i(x)]$ is an admissible weak variation for which the terminal equations of variation (9) are satisfied. There is a family $y_i = y_i(x, b)$, $x_1(b) \leqq x \leqq x_2(b)$ of admissible curves containing E_{12} for $b = 0$ and having $[\xi_1, \xi_2, \eta_i(x)]$ for variations along E_{12}. The corresponding second variation vector

$$w_i = d^2\rho_i(b)/db^2 \,|_{b=0} \qquad (i = 0, \cdots, p)$$

belongs to K_+ and hence satisfies (86). If we recall that $F = \lambda_0 f(x, y, y')$ whenever (x, y, y') is admissible, this implies that at $b = 0$ the inequality

$$(87) \quad \frac{d^2}{db^2}\left\{ \int_{x_1(b)}^{x_2(b)} F(x, y(x,b), y'(x,b))\,dx \right. \\ \left. + d_\mu \psi_\mu(x_1(b), y(x_1(b), b), x_2(b), y(x_2(b), b)) \right\} \geqq 0$$

is satisfied.

In MLP, we observed that when the rows of the matrix (69) of MLP (or of the matrix (70) of MLP) are multiplied by $1, d_1, \cdots, d_p$, respectively, and added, the sum is a row of zeros. When we expand (87), this remark shows that the coefficients of x''_s, y_{ibb} and y'_{ib} all vanish. Hence (87) takes the form

$$(88) \qquad Q[\xi_1, \eta(x_1), \xi_2, \eta(x_2)] + \int_{x_1}^{x_2} 2\omega(x, \eta, \eta')\,dx \geqq 0,$$

where

$$(89) \qquad 2\omega(x, \eta, \eta') = F_{y_i y_j}(x, y, y')\eta_i\eta_j + 2F_{y_i y'_j}(x, y, y')\eta_i\eta'_j \\ + F_{y'_i y'_j}(x, y, y')\eta'_i\eta'_j$$

and

$$(90) \qquad Q[\xi_1, \eta(x_1), \xi_2, \eta(x_2)] = [F_x(x_2, y(x_2), y'_i(x_2)) \\ + y'_i(x_2)F_{y_i}(x_2, y(x_2), y'(x_2))]\xi_2^2 \\ + 2F_{y_i}(x_2, y(x_2), y'(x_2))\xi_2\eta_i(x_2) \\ - [F_x(x, y(x_1), y'(x_1)) + y'_i(x_1)F_{y_i}(x_1, y(x_1), y'(x_1))]\xi_1^2 \\ - 2F_{y_i}(x_1, y(x_1), y'(x_1))\xi_1\eta_i(x_1) \\ + d_\mu Q_\mu[\xi_1, y'(x_1)\xi_1 + \eta(x_1), \xi_2, y'(x_2)\xi_2 + \eta(x_2)],$$

Q_μ being the quadratic form whose coefficients are the second-order partial derivatives of $\psi_\mu(x_1, y_1, x_2, y_2)$.

The inequality (88) is the basic form from which the various forms of the Jacobi (or Mayer) condition are derived. Since these further developments are well known and apply here without alteration, we proceed no further.

Our final result is then the following theorem.

THEOREM I. *For every minimizing arc for the problem of Lagrange with variable end-points there exists a set of multipliers $\lambda_0 \geqq 0$, $\lambda_1(x), \cdots, \lambda_m(x)$ such that for the function*

$$F(x, y, y', \lambda) = \lambda_0 f + \lambda_a \phi_a$$

the Du Bois–Reymond relation, the transversality condition and the Weierstrass and Clebsch conditions all hold; the sum $\lambda_0 + \Sigma \mid \lambda_a(x) \mid$ is not identically zero, and the $\lambda_a(x)$ are continuous except possibly at the corners of E_{12}. If the order of anormality of the minimizing arc is at most 1 *and hypothesis* (8) *is satisfied, we may choose the multipliers so that in addition to the preceding conclusions, the inequality* (88) *is satisfied for all admissible variations $[\xi_1, \xi_2, \eta_i(x)]$ which satisfy the terminal equations of variation* (9).

If in particular $m = 1$, so that there is only a single differential equation $\phi(x, y, y')$, the order of anormality cannot exceed 1. So we have the following theorem.

THEOREM II. *If $E_{12}: y_i = y_i(x)$, $x_1 \leqq x \leqq x_2$ is a minimizing curve for a Lagrange problem in which there is a single side-equation $\phi(x, y, y') = 0$ to be satisfied, and hypothesis* (8) *holds, then there is a non-negative constant λ_0 and a function $\lambda(x)$, continuous save perhaps at values of x defining corners of E_{12}, such that for $F(x, y, y', \lambda) \equiv \lambda_0 f + \lambda\phi$ the Du Bois–Reymond equations, the transversality conditions, and the necessary conditions of Weierstrass, Clebsch and Jacobi–Mayer are satisfied. Moreover, $\lambda_0 + \mid \lambda(x) \mid$ is not identically zero.*

9. Remark concerning the hypotheses. It might seem that the restriction on the order of anormality of E_{12} is dictated by the method of proof, and is not really essential for the validity of the conclusion. A simple example shows that this is not so.

In (x, y_1, y_2, y_3)-space consider the isoperimetric problem of minimizing

$$I[y] = \int_0^1 [y'_i y'_i - 4y_1 y'_1 - 4y_2 y'_2]\, dx$$

subject to the end-conditions

$$y_i(0) = 0, \qquad (i = 1, 2, 3)$$

and the isoperimetric conditions

$$\int_0^1 [y_1 y'_1 - y_2 y'_2]\, dx = 0,$$

$$\int_0^1 [y_1 y'_2 + y_2 y'_1]\, dx = 0.$$

These isoperimetric conditions are equivalent to

$$[y_1(1)]^2 - [y_2(1)]^2 = 0, \qquad y_1(1)y_2(1) = 0,$$

and the only solution is $y_1(1) = y_2(1) = 0$. Subject to these conditions we have

$$I[y] = \int_0^1 y'_i y'_i dx,$$

and the unique minimizing curve is $y_i(x) = 0$, $0 \leqq x \leqq 1$.

As usual, we transform our problem into a Lagrange problem by introducing two new variables y_4, y_5 together with the differential equations

$$y'_4 + y_1 y'_1 - y_2 y'_2 = 0,$$
$$y'_5 + y_1 y'_2 + y_2 y'_1 = 0,$$

and the end-conditions

$$y_4(0) = y_5(0) = y_4(1) = y_5(1) = 0.$$

If p, q and r are arbitrary real numbers, the functions

$$\eta_1 = px, \quad \eta_2 = qx, \quad \eta_3 = rx, \quad \eta_4 \equiv \eta_5 \equiv 0$$

are admissible variations and satisfy the terminal equations of variation with $\xi_1 = \xi_2 = 0$. Let $\lambda_0, \lambda_1(x), \lambda_2(x)$ be multipliers with which the Du Bois-Reymond equations hold; then, as always in isoperimetric problems, λ_1 and λ_2 are constants. The quadratic form Q is easily seen to vanish identically. So the inequality (88) reduces to

$$\int_0^1 \{2\lambda_0[p^2 + q^2 + r^2 - 4p^2 x - 4q^2 x] + 2\lambda_1[p^2 x - q^2 x] + 4\lambda_2[pqx]\}dx \geqq 0,$$

or

$$(\lambda_1 - \lambda_0)p^2 + (-\lambda_1 - \lambda_0)q^2 + \lambda_0 r^2 + \lambda_2 pq \geqq 0.$$

It is easily seen that this cannot hold for all p, q, r unless $\lambda_0 = \lambda_1 = \lambda_2 = 0$, contradicting their choice as multipliers.

UNIVERSITY OF VIRGINIA.

CERTAIN CONSEQUENCES OF THE JORDAN CURVE THEOREM.*

By F. Burton Jones.

Relatively little[1] is known of the nature of locally connected spaces in which the Jordan curve theorem holds true, except for spaces which possess other strong properties, local compactness in particular. It is the principal object of this paper to investigate the structure and properties of a locally connected, complete Moore space in which the Jordan curve theorem holds true and to show how similar such spaces are to certain subsets of a plane.[2] Certain modified forms of the Janiszewski separation theorems are shown to hold true.

Let S denote a nondegenerate locally connected complete Moore space such that (1) S contains no cut point[3] and (2) the Jordan curve theorem holds true in S.[4]

Since any two points of a connected domain are the extremities of an arc lying in the domain,[5] the following theorem is evident.

Theorem 1. *If M, the set of all non-endpoints of S, be considered as a space, then* (1) *M is a nondegenerate Moore space,* (2) *M is locally connected,* (3) *M contains no cut point of itself,* (4) *the Jordan curve theorem holds true*

* Received August 20, 1940; Presented to the American Mathematical Society, December 29, 1939.

[1] R. L. Moore, *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, XIII (1932), New York, Chapter III. This book will be referred to as *Foundations*; F. B. Jones, "Certain equivalences and subsets of a plane," *Duke Mathematical Journal*, vol. 5 (1939), pp. 133-145; F. B. Jones, "Concerning the boundary of a complementary domain of a continuous curve," *Bulletin of the American Mathematical Society*, vol. 45 (1930), pp. 428-435.

[2] Particularly those subsets of the plane used as examples in my paper, "Certain equivalences and subsets of a plane," *loc. cit.*[1]

A *Moore space* is a space satisfying Axioms 0 and the first three parts of Axiom 1 of *Foundations*. This includes the class of metric spaces.

A *complete* Moore space is a space satisfying Axioms 0 and 1 of *Foundations*. This includes the class of complete metric spaces.

[3] This is assumed mainly as a matter of convenience. The effect of omitting it may be easily ascertained from the results in my paper, "Almost cyclic elements and simple links of a continuous curve," *Bulletin of the American Mathematical Society*, vol. 46 (1940), pp. 775-783.

[4] That is, S satisfies Axioms 0 — 4 of *Foundations*.

[5] Theorem 1 of Chapter II of *Foundations*.

in M, (5) any two points of a connected open subset D of M may be joined by an arc lying wholly in D, and (6) M contains no endpoints of itself.

EXAMPLE. *There exists a nondegenerate locally connected complete Moore space S, which contains no cut point and in which the Jordan curve theorem holds true, such that the set M of all non-endpoints of S is not a complete Moore space.*

LEMMA. *In a complete Moore space, every perfect set contains a compact perfect set.*

Proof. Let H denote a perfect set. There exists a region R_{11} of G_1 of Axiom 1 of *Foundations* containing a point of H. By part (3) of Axiom 1, there exist two regions R_{21} and R_{22} each of which contains points of H such that (1) $\bar{R}_{21} \cdot \bar{R}_{22} = 0$ and (2) R_{11} contains both $\bar{R}_{21}$ and $\bar{R}_{22}$. This process may be continued so that for each integer n, there exist regions $R_{n1}, R_{n2}, R_{n3}, \cdots, R_{n2^{n-1}}$ of G_n each containing points of H such that (1) no point is common to the closure of any two of them and (2) for each i, $i = 1, 2, 3, \cdots, 2^{n-1}$, R_{ni} contains both $\bar{R}_{n+1\ 2i-1}$ and $\bar{R}_{n+1\ 2i}$. With the help of Theorem 79 on page 57 of *Foundations*, it follows that $\Pi\Sigma\bar{R}_{ni}$ is both compact and perfect. Furthermore, since every point of this set is a limit point of H and H is closed, it is a subset of H. Thus the lemma is established.

Let C denote the Cantor set (obtained by removing middle-third segments) on the interval I from $(0,0)$ to $(1,0)$ of the X-axis in the number plane E. Let S denote the subset of E remaining upon the removal from E of all points (X, Y) such that either (1) $(X, 0)$ belongs to C and $0 < Y \leqq 1$, or (2) (X, Y) belongs to the left- or right-hand 1/3 of a component of $I - C$, or (3) $(X, 0)$ is a trisection point of a component of $I - C$ and $0 \leqq Y \leqq 1$, or (4) (X, Y) belongs to one of the U-shaped continua $M_{11}, M_{21}, M_{22}, \cdots$ indicated in Figure 1 but is not one of the points $P_{11}, P_{21}, P_{22}, \cdots$. It is easy to see that S is a locally connected, inner limiting subset of E. It follows from Theorem 9 on page 96 of *Foundations* that S is a nondegenerate locally connected, complete Moore space. Furthermore, it is evident that S contains no cut point of itself and that the Jordan curve theorem holds true in S. Let M denote the subspace of S consisting of all non-endpoints of S. The space M contains all the endpoints of the components of $I - C$ but no other points of C. In M this set is perfect but countable. Hence, by the preceding lemma and Theorem 16 on page 12 of *Foundations*, M is not a *complete* Moore space.

Despite the fact that the space M of Theorem 1 is not necessarily a *complete* Moore space, many theorems proved for such spaces hold true in M since the principal result in *Foundations* requiring the use of part (4) of

Axiom 1 is: *any two points of a connected domain D are the endpoints of an arc lying in D.* With the help of Theorem 1 and the arguments and theorems of the specific references the following theorem may be established.

THEOREM 2. *Let M denote the set of all non-endpoints of S. Then* (1) *M is everywhere dense in S*; (2) *if A and B are two distinct points and*

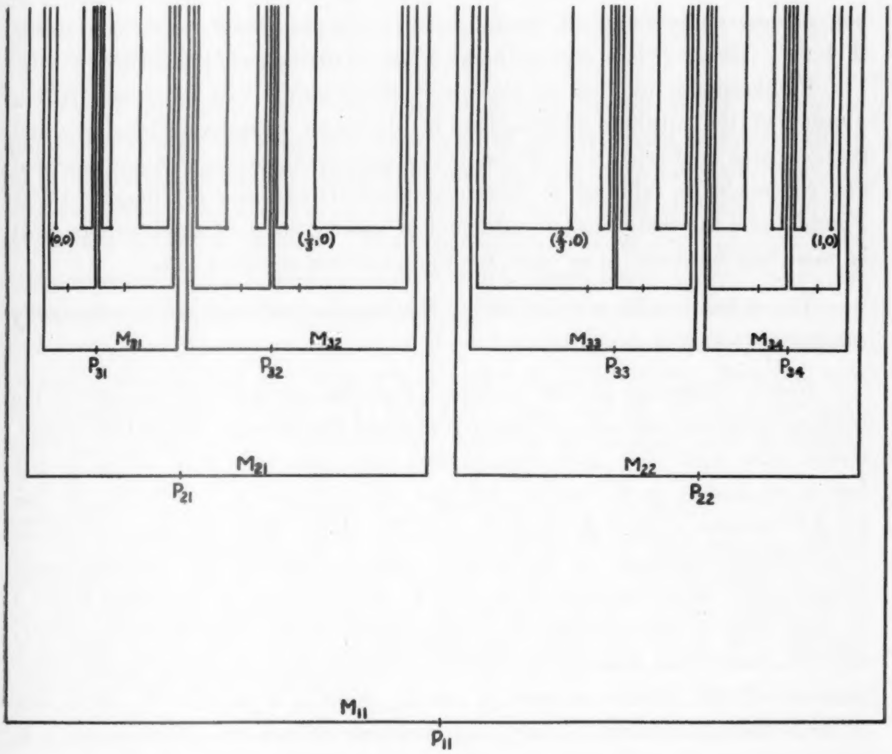

FIGURE 1.

neither of the two mutually exclusive closed and compact subsets H and K of M separates A from B, then $H+K$ does not separate A from B; (3) *if A and B are distinct points of M, there exists a simple closed curve separating A from B*; (4) *if the common part of the closed and compact subsets H and K of M is a continuum and neither H nor K separates the point A from the point B, then $H+K$ does not separate A from B*; (5) *no arc in M separates M or S*; (6) *M is cyclicly connected*; *and* (7) *if AB is an arc and J is a*

simple closed curve separating A from B, then $J + AB$ contains a simple closed curve J_1 separating A from B such that $J_1 \cdot AB$ is connected.[6]

THEOREM 3. *In order that a point P be a local endpoint of S it is necessary and sufficient that P be an endpoint of S.*

Proof. That the condition is sufficient is obvious. The condition is also necessary. For suppose that P is not a local endpoint of S. By part (1) of Theorem 2, there exists a non-endpoint A of S distinct from P. Since A is not a local endpoint of S, there exists a simple closed curve containing $A + P$.[7] Hence P is a non-endpoint of S. This is a contradiction.

Furthermore, in view of the above theorems, it can be shown that a number of the intuitive propositions of the plane concerning *abutting* and *crossing* arcs hold true in S. These definitions and theorems will not be stated and the reader is referred to Theorems 28-33 (inclusive) of Chapter IV of *Foundations* for their precise statement or proof. Some of these proofs must be modified, however, to be valid for the set of axioms used here.

THEOREM 4. *No arc in S lies in the boundary of each of three mutually exclusive connected domains.*

Proof. Suppose, on the contrary, that the arc AB lies in each of the boundaries of the mutually exclusive connected domains D_1, D_2 and D_3 respectively. Let A_1B_1 and X_1Y_1 denote arcs such that (1) $A_1B_1 - (A_1 + B_1)$ lies in D_1 and $A_1 + B_1$ lies on AB, and (2) $X_1Y_1 - X_1$ lies in D_1, X_1 lies on AB between A_1 and B_1, and $Y_1 = X_1Y_1 \cdot A_1B_1$. If J_1 denotes the simple closed curve contained in $AB + A_1B_1$, and I_1 denotes the complementary domain of J_1 which contains $X_1Y_1 - (X_1 + Y_1)$, then by Theorems 3 and 4 on page 155 of *Foundations*, $I_1 - (I_1 \cdot X_1Y_1)$ is the sum of two mutually exclusive connected domains U and V such that the closure of neither contains all of AB. Hence neither D_2 nor D_3 contains a point of I_1. It is now evident that AB contains an arc A_1B_1 such that no two arcs abutting on A_1B_1 and lying, except for their common part with A_1B_1, in D_1 and D_2, or in D_1 and D_3, respectively, abut on A_1B_1 from the same side. In a similar manner, it can be shown that A_1B_1 contains an arc A_2B_2 such that no two arcs abutting on A_2B_2 and, lying except for their common part with A_2B_2, in D_2 and D_3 respectively, abut on A_2B_2 from the same side. This involves a contradiction.

[6] For parts (2) to (6) inclusive see my paper, "Certain equivalences, etc.," *loc. cit.*[1]. Part (7) may be established by a slight modification of the argument for Theorem 10 of my paper, "Concerning certain topologically flat spaces," *Transactions of the American Mathematical Society*, vol. 42 (1937), pp. 53-93.

[7] See Theorem A′ in the footnote on page 56 of my paper on topologically flat spaces, *loc. cit.*[6]

Definition. Suppose that I and E are two complementary domains of an arc AB. Let T_I and T_E denote two arcs each having one endpoint in $AB-(A+B)$ but lying except for these two points in I and E respectively. Then I and E are said to *lie on the same side* or *opposite sides of* AB according as *every* two such arcs T_I and T_E abut on AB from the same side or opposite sides.

THEOREM 5. *If AB is an arc which separates S but which contains no proper interval separating S, then $S-AB$ is the sum of two connected domains which lie on opposite sides of AB.*

Proof. Suppose that I is a complementary domain of AB. Since no proper interval of AB separates space, both A and B belong to the boundary of I. Furthermore, suppose that X_1Y_1 and X_2Y_2 are two mutually exclusive arcs such that $(X_1Y_1-Y_1)+(X_2Y_2-Y_2)$ is a subset of I and Y_1+Y_2 is a subset of $AB-(A+B)$. Then there exists in I an arc X_1X_2 and there exists in $AB+X_1Y_1+X_1X_2+X_2Y_2$ a simple closed curve J such that $J\cdot AB$ is a connected subset of $AB-(A+B)$. Since J is a subset of $I+AB$, any complementary domain of AB other than I must lie in one of the complementary domains of J and have both A and B in its boundary. Hence J does not separate A from B, and X_1Y_1 and X_2Y_2 must abut on AB from the same side. Let I_1 denote the complementary domain of J which fails to contain $A+B$. Since I_1 contains no point of AB but has limit points in I, I_1 is a subset of I. Hence, in addition to A and B, every point between Y_1 and Y_2 on AB is in the boundary of I. With the help of Theorem 2 on page 89 of *Foundations*, it follows that AB is the boundary of I. By Theorem 4, there are only two complementary domains of AB and since any two arcs such as X_1Y_1 and X_2Y_2 above abut on AB from the same side, these two domains must lie on opposite sides of AB.[8]

THEOREM 6. *If A is an endpoint of S and AB is an arc separating S, then AB contains one and only one point X such that the interval AX of AB separates S but contains no proper interval separating S.*

Proof. If B is not the point X, then AB contains a point B_1 distinct from B such that the interval AB_1 of AB separates S. By Theorem 2, part (5), every interval of AB_1 which separates S contains A. Hence every component of $S-AB_1$ has A in its boundary. Since A is not a cut point of S, each component of $S-AB_1$ has a boundary point in AB_1 distinct from A. Let D denote a complementary domain of AB_1. Suppose that X, Y and Z

[8] The reader can easily see that given any arc T_1 which abuts on AB, there exists an arc T_2 such that T_1 and T_2 abut on AB from opposite sides.

are distinct points lying on AB_1 in that order, such that X and Z belong to the boundary of D. There exist two arcs OX_1 and OZ_1 such that (1) $(OX_1 - X_1) + (OZ_1 - Z_1)$ is a subset of D, (2) X_1 and Z_1 belong to $AB_1 - (A + B_1)$ in the order X_1, Y, Z_1, and (3) OX_1 and OZ_1 have only O in common. If OX_1 and OZ_1 abut on AB_1 from the same side, then one of the complementary domains of $OX_1 + X_1Z_1$ (of AB_1) $+ OZ_1$ contains neither A nor B_1 and consequently is a subset of D. So Y belongs to the boundary of D. On the other hand, if OX_1 and OZ_1 abut on AB_1 from opposite sides and I denotes the complementary domain of $OX_1 + X_1Z_1$ (of AB_1) $+ OZ_1$ which contains B_1, then (1) the domain $I - (I \cdot AB_1)$ is, by parts (4) and (5) of Theorem 2, connected, (2) it is a subset of D, and (3) its boundary contains Y. So again Y belongs to the boundary of D. Since the boundary of D contains every point of AB_1 lying between any two points of its boundary, the boundary of D is an interval of AB_1 which contains A. It follows from Theorem 4 that there are two complementary domains of AB_1. The boundary of one of these domains is a subset of the boundary of the other and the endpoint of its boundary distinct from A is the point X of Theorem 6.

Theorem 7. *If AB is an arc containing intervals AX and BY such that each separates S but no proper interval of either separates S, then AX and BY have at most one point in common or $AX = BY = AB$.*

Theorem 7 follows, by an indirect argument, immediately from Theorems 4 and 5.

Theorem 8. *Suppose that AB is an arc. Then* (1) *if neither A nor B is an endpoint of S, AB has one complementary domain,* (2) *if only one of the points A and B is an endpoint of S, AB has at most two complementary domains, and* (3) *if both A and B are endpoints of S, AB has at most three complementary domains.*

Theorem 9. *If no simple closed curve separates the point A from the point B, then every arc from A to B separates space.*[9]

Theorem 10. *If the arc AB separates space but no proper interval of AB separates space, then no simple closed curve separates A from B.*

Proof. Suppose that there exists a simple closed curve J separating A from B. By Theorem 2, part (7), $J + AB$ contains a simple closed curve J_1 separating A from B such that $J_1 \cdot AB$ is connected. One of the two com-

[9] See footnote 6 on page 135 of my paper on "Certain equivalences, etc.," *loc. cit.*[1]

plementary domains of AB contains $J_1 - J_1 \cdot AB$. But this contradicts Theorem 5, since it follows from Theorem 32 on page 201 of *Foundations* that $J_1 - J_1 \cdot AB$ contains arc segments whose closures are arcs which abut on AB from opposite sides.

THEOREM 11. *If A and B are distinct points and some arc from A to B separates space but contains no proper interval separating space, then every arc from A to B separates space.*

THEOREM 12. *If A is an endpoint of S, there is at most one non-endpoint X of S such that every arc from A to X separates space.*

Proof. Suppose, on the contrary, that there are two non-endpoints X and Y of S such that every arc from A to X or from A to Y separates space. Let AX denote an arc from A to X in $S - Y$. By Theorem 6, AX contains a point X_1 such that the interval AX_1 of AX separates space but contains no proper interval separating space. There exists an arc AY in $S - X_1$. Let Y_1 denote a point of AY such that the interval AY_1 of AY separates space but contains no proper interval separating space. Since the point Y_1 is distinct from X_1, it follows from Theorem 11 that Y_1 does not belong to AX_1. By Theorem 2, part (3), there exists a simple closed curve J which separates X_1 from Y_1. But A, being an endpoint of S, does not belong to J. Hence, either J separates A from X_1 or J separates A from Y_1. This is contrary to Theorem 10.

THEOREM 13. *If A is an endpoint of S, there is at most one non-endpoint X of S such that no simple closed curve separates A from X.*

Theorem 13 follows from Theorems 9 and 12.

THEOREM 14. *If A and B are distinct points and every arc from A to B separates space, then no simple closed curve separates A from B.*

Proof. Two cases arise depending upon whether or not both of the points A and B are endpoints of S.

Case I. Suppose that B is a non-endpoint of S. It follows from Theorem 2, part (5), that A is an endpoint of S. If AB is an arc from A to B, it follows from Theorem 6 that AB contains one and only one point X such that the interval AX of AB separates space but contains no proper interval separating space. By Theorem 11, every arc from A to X separates space. Hence, by Theorem 12, X is B. So no proper interval of AB separates space. By Theorem 10, no simple closed curve separates A from B.

Case II. Suppose that each of the points A and B is an endpoint of S.

By Theorem 12, there is at most one non-endpoint X of S such that every arc from A to X separates S. Likewise, there is at most one non-endpoint Y of S such that every arc from B to Y separates space. By Theorem 2, part (2), $S-(X+Y)$ is connected. Let AB denote an arc from A to B in $S-(X+Y)$. It follows from Theorems 6 and 11 that no proper interval of AB separates space. Hence, by Theorem 10, no simple closed curve separates A from B.

THEOREM 15. *If A and B are two distinct points, then in order that every arc from A to B separate space, it is necessary and sufficient that no simple closed curve separate A from B.*

THEOREM 16. *If A, B and C are three distinct points, A is an endpoint of S and every arc from A to either B or C separates space, then every arc from B to C separates space.*

Theorem 16 follows immediately, by an indirect argument, from Theorem 15.

THEOREM 17. *Suppose that H and K are two mutually exclusive closed and compact point sets such that if X and Y are any two points of H and K, respectively, there exists a simple closed curve separating X from Y. Then if A and B are two points belonging to H and K respectively, there exists a simple closed curve which separates A from B and contains no point of $H+K$.*

Proof. Let X denote a point of H. For each point Y of K, let I_Y denote a simple domain which contains Y such that $\bar{I}_Y$ does not contain X. There exists a finite collection Δ_1 of these domains covering K. By Theorems 2, part (3), and 13, there is at most one non-endpoint Z of S such that no simple closed curve separates X from Z. From each element of Δ_1 select a point distinct from Z, join each pair of these points by an arc in $S-(X+Z)$, and let L denote the sum of these arcs. There exists a finite collection Δ_2 of simple domains covering L in the same way that Δ_1 covers K. Evidently $\Delta_1^*+\Delta_2^*$ is connected, and contains K.[10] It follows from Theorem 13 on page 166 of *Foundations*, that there exists a simple closed curve J_X which separates X from $\Delta_1^*+\Delta_2^*$ and, hence, from K. So for each point X of H there exists a simple domain I_X containing X such that $\bar{I}_X$ contains no point of K. There exists a finite collection G of these domains such that G covers H. Let D denote the component of G^* which contains A. By Theorem 13 on page 166 of *Foundations* there exists a simple closed curve C which sepa-

[10] The notation Δ^* means the sum of the elements of Δ.

rates B from D and which is a subset of the boundary of D. Evidently C separates A from B and contains no point of $H + K$.[11]

Notation. If P is a point, then N_P is used to denote P together with the set of all points X such that no simple closed curve separates X from P.

THEOREM 18. *If P is a point, N_P is closed.*

Proof. Suppose that Y is a limit point of N_P. If some simple closed curve J separates Y from P, then it is clear that the complementary domain of J which contains Y contains points of N_P contrary to the definition of N_P. Hence Y belongs to N_P.

THEOREM 19. *If P is a point, N_P contains at most one non-endpoint of S.*

Proof. Suppose, on the contrary, that N_P contains two non-endpoints X and Y of S. It follows from Theorem 2, part (3), that neither X nor Y is P and, in fact, that P is an endpoint of S. By Theorem 15, every arc from P to either X or Y separates space. This contradicts Theorem 12.

THEOREM 20. *Suppose that O and P are distinct points such that N_O and N_P have at least two points in common. Then* (1) *one of the two points O and P is an endpoint of S,* (2) *if O is an endpoint of S, N_O is a subset of N_P,* (3) *if both O and P are endpoints of S, then N_O and N_P are identical, and* (4) *if N_O is a proper subset of N_P, then P is a non-endpoint of S.*

Proof. Suppose that there exists a simple closed curve J separating O from P. Let I_O and I_P denote the complementary domains of J containing O and P respectively. It follows from the definition of the sets N_O and N_P that $\bar{I}_O$ and $\bar{I}_P$ contain N_O and N_P respectively. By Theorem 19, J contains at most one point of each of the sets N_O and N_P. Being a subset of J, $N_O \cdot N_P$ contains at most one point, contrary to hypothesis. Hence no simple closed curve separates O from P. It follows from Theorem 2, part (3), that one of the points O and P is an endpoint of S, and this establishes part (1) of Theorem 20. But if O is an endpoint, no simple closed curve separates a point X from P without also separating X from O. Hence if O is an endpoint, N_P contains N_O. This establishes part (2). Parts (3) and (4) follow from part (2).

THEOREM 21. *If P is a point, N_P contains no nondegenerate compact continuum.*

Proof. If N_P contains a nondegenerate compact continuum, then it

[11] The last part of this argument is part of the argument to prove Theorem 9 on page 184 of *Foundations.*

follows with the help of Theorem 19 that N_P contains a nondegenerate compact continuum K containing only endpoints of S. Let X denote a point of K. By Theorem 3, X is a local endpoint of S. Hence there exists a sequence $V_1, V_2, V_3, \cdots$ of connected domains such that (1) K is not a subset of V_1, (2) $V_1 \supset V_2$, $V_2 \supset V_3$, $V_3 \supset V_4, \cdots$, (3) for each integer i, $i > 1$, the boundary of V_i with respect to V_1 contains only one point, X_i, and (4) $\Pi V_i = X$. Let R denote a region which contains X and lies together with its boundary, β, in V_1. Then $\beta \cdot K$ is closed, compact, and non-vacuous. Let C denote the component of $\bar{R} \cdot K$ which contains X. Since K is a compact continuum, C contains a point of β. Since each of the points $X_2, X_3, X_4, \cdots$ is a cut point of V_1, no one of them is an endpoint of S, and hence no one of them belongs to C. Consequently, C is, for each integer i, a subset of V_i, and ΠV_i contains C. This is a contradiction.

THEOREM 22. *If P is a point, $S - N_P$ is connected.*

Proof. Suppose, on the contrary, that N_P separates the point A from the point B. By Theorem 19, N_P contains at most one non-endpoint O. There exists an arc AB in $S - O$. But $AB - (A + B)$ must contain a point of N_P distinct from O. This is a contradiction.

THEOREM 23. *Suppose that P is an endpoint of S belonging to a nondegenerate continuum K which is locally peripherally compact at P. If R is a region containing P, then R contains an open subset r of K containing P such that* (1) *the boundary of r with respect to K contains only one point X,* (2) *$r + X$ is connected and* (3) *X is a non-endpoint of S.*

Proof. By Theorem 3, P is a local endpoint of S. Hence there exists a sequence $V_1, V_2, V_3, \cdots$ of connected domains such that (1) $R \supset V_1$, $V_1 \supset V_2$, $V_2 \supset V_3, \cdots$, (2) for each integer i, $i > 1$, the boundary of V_i with respect to V_1 contains only one point, X_i, and (3) $\Pi V_i = P$. Let D denote a domain containing P and lying together with its boundary in V_1 such that, if β is the boundary of D, $\beta \cdot K$ is compact. There exists an integer j such that $\bar{V}_j \cdot \beta \cdot K = 0$, X_j is in D, and K is not a subset of $D \cdot V_j$. Hence, K contains a boundary point of $D \cdot V_j$. Let γ denote the boundary of V_j. Any point of K in the boundary of $D \cdot V_j$ belongs either to $\beta \cdot V_j$, to $\gamma \cdot D$, or to $\beta \cdot \gamma$. But since the first and last contain no point of K, and $\gamma \cdot D$ is X_j, it follows that the only point of K in the boundary of $D \cdot V_j$ is X_j. Let r denote $D \cdot V_j \cdot K$ and let X denote X_j. Evidently r is an open subset of K containing P whose boundary with respect to K is the one point X. Since X is a cut point of V_1, X is a non-endpoint of S. If $r + X$ is not connected, it is the sum of two mutually exclusive closed sets, w and q, such

that w contains X. Then $(K - r) + w$ and q are two mutually exclusive closed sets whose sum is K. This is a contradiction.

THEOREM 24. *Suppose that P is a non-endpoint of S belonging to a continuum K which is locally compact at P. If R is a region containing P, then R contains a compact open subset r of K which contains P and whose boundary with respect to K contains no endpoint of S.*

Proof. There exists in R a domain D containing P such that $\bar{D} \cdot K$ is compact. It follows from Theorems 18 and 21 that $\bar{D} \cdot K \cdot N_P$ is totally disconnected. Hence $\bar{D} \cdot K$ contains an open subset w of K which contains P and whose boundary with respect to K contains no point of N_P. Let β denote the boundary of w with respect to K. By Theorems 18 and 22, $S - N_P$ is a connected domain. It follows from Theorem 9 on page 96 of *Foundations* and a theorem of Miss H. C. Miller[12] that β is a subset of a compact continuum H lying in $S - N_P$. By Theorem 17, there exists a simple closed curve J separating P from H. Let I denote the complementary domain of J which contains P. Then $I \cdot w$ is an open subset r of K which contains P and whose boundary with respect to K, being a subset of J, contains no endpoint of S.

THEOREM 25. *No finite set of points separates space.*

Proof. Suppose, on the contrary, that a finite set H separates the point A from the point B. It follows from Theorem 2, part (2), that the set H_1 of non-endpoints of S in H does not separate A from B. Let AB denote an arc from A to B lying in $S - H_1$. Since AB must contain a point of H, AB contains a point X of H which is an endpoint of S. But X can be neither A nor B, which is a contradiction.

THEOREM 26. *If H and K are two continua such that $H \cdot K$ is not connected and the boundary of $H \cdot K$ with respect to K is a subset of a compact open subset of K, then $H + K$ separates S.*[13]

Proof. Since K is connected and $H \cdot K$ is not connected, the boundary

[12] *In a locally connected complete Moore space, any closed and compact subset of a domain is itself a subset of a compact continuum in the domain.* Miss H. C. Miller, "A theorem concerning closed and compact point sets which lie in connected domains," *Bulletin of the American Mathematical Society*, vol. 46 (1940), p. 848.

[13] Theorem 26 is a modification of the well-known Janiszewski theorem for the plane. Cf. Theorem 20 on page 192 of *Foundations*. If in Theorem 26 the words "the boundary of $H \cdot K$ is a subset of a compact open subset of K" are replaced by "the boundary of $H \cdot K$ is compact" or "$H \cdot K$ is compact" or even "$H \cdot K$ is compact and both $H - H \cdot K$ and $K - H \cdot K$ are connected," the resulting proposition is false.

of $H \cdot K$ with respect to K is the sum of two mutually exclusive non-vacuous closed sets M and N. There exist two compact and mutually exclusive open subsets d_M and d_N of K which contain M and N respectively. It follows from Theorems 23 and 24 that there exist two open subsets r_M and r_N of K containing M and N respectively such that (a) $d_M \supset \bar{r}_M$ and $d_N \supset \bar{r}_N$ and (b) neither the boundary of r_M with respect to K nor the boundary of r_N with respect of K contains an endpoint of S. Let Δ_M denote the collection of all components of r_M which contain a point of M and let Δ_N denote the collection of all components of r_N which contain a point of N. If P is a boundary point of $\overline{\Delta_M{}^*}$ with respect to K, then either (1) P belongs to the boundary of r_M with respect to K or (2) P belongs to r_M and is the sequential limit point of a sequence α of points of r_M no two of which belong to the same component of r_M. In case (1), P is a non-endpoint of S. In case (2), suppose that P is an endpoint of S. It follows from Theorem 23, that there exists an open subset r_P of K which contains P and whose boundary with respect to K is a single point X such that $r_P + X$ is a connected subset of r_M. The set r_P contains two points of α. Since these points belong to distinct components of r_M, this is a contradiction. Hence in case (2), P is a non-endpoint of S. Likewise, the boundary of $\overline{\Delta_N{}^*}$ with respect to K contains no endpoint of S. Let H_1 denote $H + \overline{\Delta_M{}^*} + \overline{\Delta_N{}^*}$. Clearly H_1 is a continuum. Let K_1 denote the closure of $K - (\overline{\Delta_M{}^*} + \overline{\Delta_N{}^*})$ and let M_1 and N_1 denote $H_1 \cdot K_1 \cdot d_M$ and $H_1 \cdot K_1 \cdot d_N$ respectively. The sets M_1 and N_1 are closed, mutually exclusive and non-vacuous, and $H_1 \cdot K_1 = M_1 + N_1$. Furthermore, neither M_1 nor N_1 contains an endpoint of S. Suppose that K_1 is the sum of two mutually exclusive closed sets M_2 and N_2 which contain M_1 and N_1 respectively. Then $M_2 + \overline{\Delta_M{}^*}$ and $N_2 + \overline{\Delta_N{}^*}$ are two mutually exclusive closed sets whose sum is K. Since K is a continuum, this is a contradiction. Hence K_1 is not the sum of two mutually exclusive closed sets containing M_1 and N_1 respectively. It follows indirectly from the argument for Theorem 35 on page 21 of *Foundations* that there exist a point A of M_1 and a point B of N_1 such that K_1 is not the sum of two mutually exclusive closed sets containing A and B respectively. By Theorem 17, there exists a simple closed curve J which separates A from B and contains no point of $M_1 + N_1$. Let H_2 and K_2 denote $H_1 \cdot J$ and $K_1 \cdot J$ respectively. Neither H_2 nor K_2 is vacuous since neither H_1 nor K_1 is the sum of two mutually exclusive closed sets containing A and B respectively. The proof of Theorem 26 may now be completed by making slight notational changes in the argument on page 192 of *Foundations*.[14]

[14] Moore attributes this argument to Kuratowski—presumably the argument on pages 222 and 223 of the paper, "Sur la séparation d'ensembles situés sur le plan," *Fundamenta Mathematicae*, vol. 12 (1928), pp. 214-239.

THEOREM 27. *If A and B are distinct points, no simple closed curve separates A from B, and K is a continuum containing $A + B$ which is locally compact at A, then K separates space.*

Proof. By Theorems 19 and 20, $N_A + N_B$ contains at most one non-endpoint X. It follows from Theorems 23 and 24 that there exists an open subset r of K such that (1) r contains A, (2) β, the boundary of r with respect to K, is compact and contains no endpoint of S, and (3) $\bar{r}$ contains neither B nor X. Since β is closed and compact and separates A from B in K, β contains a closed set T which separates A from B in K and contains no closed proper subset which separates A from B in K. The set $K - T$ is the sum of two mutually separate sets K_A and K_B containing A and B respectively, and T is irreducible with respect to separating A from B in K. By Theorem 92 on page 66 of *Foundations* both $K_A + T$ and $K_B + T$ are connected. Let Y denote a point of T. By Theorem 17, there exists a simple closed curve J which separates Y from A and contains no point of $T + (A + B)$. Since J does not separate A from B, J separates Y from B. Let K_1 and K_2 denote $J \cdot K_A$ and $J \cdot K_B$ respectively. Each of these sets is closed and non-vacuous, and they are mutually exclusive. There exist two mutually exclusive closed sets N_A and N_B such that each is the sum of a finite number of mutually exclusive arcs of J and each point of K_A and K_B is a non-endpoint of a component of N_A or N_B respectively. There are only a finite number of components in $J - (N_A + N_B)$. If K does not separate space, select one point from each component of $J - (N_A + N_B)$, join each pair of these points by an arc in $S - K$, and let V denote the closure of $J - (N_A + N_B)$ together with the sum of these arcs. The set V contains no endpoint of S. Let D denote the complementary domain of $J + V$ which contains A. It follows from Theorem 2, part (2), that the outer boundary J_1 of D with respect to Y is connected. By Axiom 3, J_1 is not degenerate. Since $J + V$ is a compact continuous curve, J_1 is either an arc or a simple closed curve.[15] By Theorem 2, part (3), J_1 is not an arc. Hence J_1 is a simple closed curve. Since no point of $J + V$ is a cut point of $J + V$, J_1 is also the boundary of D.[16] Hence D contains B. The set J_1 contains no point of the complementary domain of J which contains Y. But $N_B + V$ does not separate Y from A; since D contains $A + B$, $N_A + V$ does not separate Y from A; and $(N_A + V) \cdot (N_B + V)$ is the connected set V. By Theorem 2, part (4), $N_A + N_B + V$ does not separate Y from A. This is a contradiction, for $N_A + N_B + V$ contains the simple closed curve J.

---

[15] This follows from Theorem 8 of my paper, "Concerning the boundary of a complementary domain of a continuous curve," *loc. cit.*[1]

[16] *Ibid.*, Theorem 12.

THEOREM 28. *Suppose that J is the boundary of a simple domain I and that H is a connected set and K is a continuum, each lying in $I + J$. Then if H and K are mutually exclusive, no two points of $H \cdot J$ separate any two points of $K \cdot J$ from each other on J.*[17]

Proof. Suppose, on the contrary, that the points A and B of $H \cdot J$ separate the points C and D of $K \cdot J$ from each other on J. By Theorem 18 on page 169 of *Foundations*, $\bar{I}$ is a continuous curve. There exists a collection Δ of connected open subsets of $\bar{I}$ such that (1) Δ covers H, (2) every element of Δ contains a point of H, and (3) no element of Δ contains a point of K. Since Δ^* is a connected open subset of $\bar{I}$, there exists an arc CD lying in Δ^*. In a similar manner, there exists a collection G of connected open subsets of $\bar{I}$ such that G covers K, every element of G contains a point of K, and no element of G contains a point of CD. In G^* there exists an arc AB. Since AB and CD have no point in common, this contradicts Theorem 17 on page 167 of *Foundations.*

The example in Figure 9 on page 167 of *Foundations* shows that (even in a plane) if the assumption that K is closed is omitted from the hypothesis of Theorem 28, the resulting proposition is false. This example also makes the truth of Theorem 28 more surprising for the space considered here than for a plane.

Unsolved problem. In connection with Theorems 18 and 21, I have been unable to show that *N_P is totally disconnected.* Speaking roughly, it *is* if the space contains no "hills." This is the case in the various examples in my paper "Certain equivalences and subsets of a plane."[18]

THE UNIVERSITY OF TEXAS.

---

[17] Cf. Theorem 17 on page 167 of *Foundations.* Theorem 28 has been discovered independently by Miss H. C. Miller. Miss Miller has also shown that the word "compact" may be entirely omitted from Theorem 20 on page 170 of *Foundations.*

[18] *Loc. cit.*[1]

APOSYNDETIC CONTINUA AND CERTAIN BOUNDARY PROBLEMS.* [1]

By F. Burton Jones.

As a natural generalization of the notion of *connectedness im kleinen* at a point, I define what is meant by a continuum being *aposyndetic* at a point. If a continuum is aposyndetic at each of its points, then the continuum is said to be aposyndetic. The class of aposyndetic continua includes both the class of continuous curves and the class of semi-locally connected continua (whether locally compact or not) and is identical with the class of freely decomposable continua. Furthermore, every aposyndetic continuum (in a Hausdorff space) which is locally peripherally bicompact is semi-locally connected. With this and the definitions in mind, it is easy to see that many (perhaps all) of the theorems proved by G. T. Whyburn for semi-locally connected continua [2] also hold true for aposyndetic continua. The utility of the notion of a continuum being aposyndetic *at a point* is illustrated by the following theorem for certain abstract spaces in which the Jordan curve theorem holds true: if P is a point of a continuum M such that (1) M is locally compact at P and (2) M is aposyndetic at P, then M together with all but a finite number of its complementary domains is connected im kleinen at P. As an interesting application of this theorem, it is shown that a proposition

* Received August 23, 1940.

[1] Presented to the American Mathematical Society, September 12, 1940, under the title: "Quasi-continuous curves and certain boundary problems."

[2] G. T. Whyburn defined the term, semi-locally connected, in a paper, "Semi-locally connected sets," *American Journal of Mathematics*, vol. 61 (1939), pp. 733-749. In this paper he acknowledged the similarity of "semi-local connectedness" to R. L. Wilder's notion, "local 0-avoidability" (R. L. Wilder, "Sets which satisfy certain avoidability conditions," *Casopis pro Pestovani Matematiky a Fysiky*, vol. 67 (1938), pp. 185-198). Wilder subsequently introduced a notion, "almost i-avoidable," still more closely related to "semi-local connectedness" (R. L. Wilder, "Property S_n," *American Journal of Mathematics*, vol. 61 (1939), pp. 823-832). It seems to have gone unnoticed that Whyburn had introduced years before the notion of "semi-local connectedness" calling it "local divisibility" in his paper, "The cyclic and higher connectivity of locally connected spaces," *American Journal of Mathematics*, vol. 53 (1931), pp. 427-442. I used this property (local divisibility) of Whyburn's to remove a certain kind of oscillation from a two-dimensional continuous curve in a paper, "Certain equivalences and subsets of a plane," *Duke Mathematical Journal*, vol. 5 (1939), pp. 133-145, with, unfortunately, no knowledge of, or reference to, the other terms employed for this notion.

analogous to Whyburn's generalization[3] of the Torhorst theorem for the plane holds true in these abstract spaces.

1. Aposyndetic continua. In this section, suppose that the set of all points, S, is a Hausdorff space, the fundamental and undefined terms being point and region (open set). The following definition is a rather natural extension of the notion of connectedness im kleinen at a point.[4]

Definitions. The point set M is said to be *aposyndetic* at the point P if P belongs to M and for each point X of M distinct from P there exists an open subset of M which contains P and belongs to a connected and relatively closed subset of M lying in $M - X$. A point set which is aposyndetic at each of its points is said to be *aposyndetic*.[5]

THEOREM 1. *In order that a continuum M be aposyndetic at a point P of M it is necessary and sufficient that for each point X of M distinct from P there exist a continuum lying in $M - X$ which contains an open subset of M containing P.*

THEOREM 2. *If a point set M is connected im kleinen at a point P of M, then M is aposyndetic at P.*

THEOREM 3. *A semi-locally connected continuum is an aposyndetic continuum.*[6]

Proof. Suppose that P is a point of a semi-locally connected continuum M. Let X denote a point of M distinct from P and let R denote a region

---

[3] THEOREM. *If M is a compact semi-locally connected plane continuum, the boundary of each complementary domain of M is locally connected.*

[4] For the definition of certain terms used in this paper for which a definite reference is not given, the reader is referred to Sierpinski's *Introduction to General Topology*, translated by C. Cecilia Krieger, The University of Toronto Press, Toronto, 1934, or to R. L. Moore's *Foundations of Point Set Theory*. The latter will be referred to as *Foundations*.

[5] Aposyndetic (Gk. apo = *away from*, syn = *together*, deo = *to bind*) means bound together away from. If G is a collection of subsets of a set M, to say that M is G-aposyndetic at a point P means that if g is an element of G not containing P, then there exists a connected and relatively closed subset of M lying in $M - g$ which contains an open subset of M containing P. If the point set M is M-aposyndetic at a point P, it is said to be aposyndetic at P for simplicity, when this usage causes no confusion.

[6] A point set M is said to be semi-locally connected at a point P of M provided that every open subset of M containing P contains an open subset of M containing P whose complement in M has at most a finite number of components. A point set which is semi-locally connected at each of its points is said to be semi-locally connected. See G. T. Whyburn, *loc. cit.*[2]

which contains X such that $\bar{R}$ does not contain P. Since M is semi-locally connected at X, it follows that R contains an open subset V of M which contains X such that $M - V$ is the sum of the elements of a finite collection G of mutually exclusive continua. The element of G which contains P contains an open subset of M containing P but does not contain X.

Example. The converse of neither Theorem 2 nor Theorem 3 is true, for *an aposyndetic continuum may contain a point at which it is neither connected im kleinen nor semi-locally connected.* Let S denote the subspace of the Euclidean number plane consisting of all points of the plane except those points $(0, Y)$ such that $0 < |Y| \leqq 1$. Let M denote the origin together with the graph of the equation, $y = \sin 1/x, x \neq 0$. The continuum M is aposyndetic. It is both connected im kleinen and semi-locally connected at every point except the origin where it is neither connected im kleinen nor semi-locally connected. However, upon the addition to M of the points of S whose ordinates are $\pm 1/2$, a semi-locally connected continuum M_1 is obtained which is still not connected im kleinen at the origin. But upon the addition of the x-axis to M a continuous curve M_2 is obtained which is still not semi-locally connected at the origin.

THEOREM 4. *If an aposyndetic continuum M is locally peripherally bicompact at a point P of M, then M is semi-locally connected at P.*

Proof. Let R denote a region containing P. There exists in R an open subset U of M which contains P such that β, the boundary of U with respect to M, is bicompact. For each point O of β, there exists a continuum M_O which lies in $M - P$ and contains an open subset of M containing O. Hence, there exists a finite collection G of n such continua covering β, n being an integer. Let V denote $U - (U \cdot G^*)$.[7] The set V is an open subset of M lying in U and containing P. Furthermore, there are at most n components of $M - V$. Hence, M is semi-locally connected at P.

COROLLARY. *A locally peripherally bicompact aposyndetic continuum is semi-locally connected.*

Definition. A continuum M is said to be *freely decomposable* provided that if A and B are distinct points of M then M is the sum of two continua neither of which contains both A and B.

THEOREM 5. *Every freely decomposable continuum is aposyndetic.*

---

[7] The notation G^* denotes the sum of the elements of G.

Proof. If P is a point of a freely decomposable continuum M, and X is a point of M distinct from P, then there exist two continua M_P and M_X containing P and X, respectively, neither of which contains both P and X. Since $M - M_X$ is an open subset of M which lies in M_P and contains P, it is clear that M is aposyndetic at P.

THEOREM 6. *Every aposyndetic continuum is freely decomposable.*[8]

Proof. Let A and B denote distinct points of the aposyndetic continuum M. There exist sets H, K, U, and V such that (1) U and V are open subsets of M containing A and B respectively, (2) H and K are subcontinua of M containing U and V respectively, and (3) $H \cdot V = K \cdot U = 0$. Let H_1 denote the component of $M - V$ that contains H, and let K_1 denote the component of $M - U$ that contains K. Suppose that $H_1 + K_1 \neq M$, i. e., $L = M - (H_1 + K_1) \neq 0$. Since V is a subset of K, $H_1 + \bar{L}$ is a closed subset of $M - V$ containing H_1 as a proper subset. Hence $H_1 + \bar{L}$ is the sum of two mutually exclusive closed sets H_2 and L_1, where H_2 contains the continuum H_1. Then $K_1 + L_1$ is a closed subset of $M - U$ containing K_1 as a proper subset. Consequently, $K_1 + L_1$ is the sum of two mutually exclusive closed sets K_2 and L_2, where K_2 contains K_1. Thus M is the sum of the two mutually exclusive closed sets $(H_2 + K_2)$ and L_2 which is contrary to hypothesis.

THEOREM 7. *In order that a continuum be aposyndetic it is necessary and sufficient that it be freely decomposable.*

To obtain a parallel to Theorem 7 for continuous curves it will be necessary to prove the following proposition.

THEOREM 8. *In order that the locally peripherally bicompact continuum M be a continuous curve it is necessary and sufficient that if A is a point of M and B is a subcontinuum of M not containing A, then M is the sum of two continua neither of which contains both A and a point of B.*[9]

Proof. The condition is sufficient. For let P denote a point of M lying in a region R. By Theorems 4 and 5 there exists an open subset V of M

---

[8] Both Theorem 6 and its proof were altered after refereeing. As originally stated, Theorem 6 would have applied only to locally peripherally bicompact continua.

[9] Cf. W. A. Wilson, "A property of continua equivalent to local connectivity," *Bulletin of the American Mathematical Society*, vol. 36 (1930), pp. 85-88, § 4, p. 88, and *Foundations*, Theorem 51, p. 134, and Theorem 52, p. 136. Both of these theorems follow from Theorem 8 with very little argument. Furthermore, Theorems 7 and 8 clear up two unanswered questions about decomposable continua which arise in Wilson's paper.

lying in R and containing P such that $M - V$ is the sum of the elements of a finite set G of mutually exclusive continua. Let D denote a component of $M - P$ that contains an element of G. The set D is an open subset of M having P as its only boundary point with respect to M, and, since D is strongly connected, it can be shown that there exists in D a continuum B which contains all of the continua of G in D.[10] By hypothesis M is the sum of two continua neither of which contains both P and a point of B. Let H denote that one of these two continua which contains P. Evidently $H \cdot \bar{D}$ is a continuum lying in $V \cdot \bar{D}$ which contains an open subset of $\bar{D}$ containing P. There being only a finite number of components of $M - P$ which contain an element of G, the closure M_1 of their sum is connected im kleinen at P. Furthermore, the complement (in M) of their sum is a continuum T lying in V and containing P. Let L denote the component of $V \cdot M_1$ which contains P. Then $T + L$ is the component of V which contains P. It is clear that $T + L$ contains an open subset of M containing P. Hence M is connected im kleinen at P.

The condition is also necessary. Let D denote a connected open subset of M which contains A but no point of some open subset of M containing B.[11] There exists in D an open subset V of M which contains A such that $M - V$ is the sum of the elements of a finite collection G of mutually exclusive continua. The set $\bar{D}$ together with all of the elements of G not containing a point of B is a continuum; the set B together with all of the continua of G which contain a point of B is a continuum; the sum of these two continua is M but neither contains both A and a point of B.

The preceding theorems show, to some extent, the similarity that exists between continuous curves and aposyndetic continua. The following theorem shows a similarity that exists between continua which are not continuous curves and continua which are not aposyndetic.

THEOREM 9. *If the continuum M is not aposyndetic at the point P of M, then there exists a point X of $M - P$ such that if G is a finite set of continua filling up M then at least one continuum of the set G contains both P and X.*[12]

Proof. Suppose the contrary. Let X denote a point of $M - P$. There exists a finite set G of continua filling up M such that no continuum of G

[10] G. T. Whyburn, "Semi-locally connected sets," *loc. cit.*[2], Theorems 6.1 and 6.21, p. 737.

[11] Any locally peripherally bicompact continuum is regular.

[12] Cf. Theorem 54 on page 138 of *Foundations*.

contains both P and X. Let H denote the sum of the elements of G which contain P. Obviously H is a subcontinuum of M which contains P. Let K denote the sum of the elements of G which do not contain P. The set K is closed and $M - P$ is an open subset of M containing P and lying in H. This is contrary to the hypothesis of Theorem 9.

Remarks. Many, and possibly all, of the theorems of G. T. Whyburn's paper on semi-locally connected continua[13] hold true when "aposyndetic continuum" is substituted for "semi-locally connected continuum." This is a direct consequence of Theorem 4 for the case of locally compact continua. In other cases, Whyburn's arguments may be used. Hence, for structure considerations the reader is referred to Whyburn's paper.

2. Boundary problems. In this section it is assumed that S, the set of all points, is a non-degenerate, locally connected, complete Moore space such that (1) S contains no cut point and (2) the Jordan curve theorem holds true in S. In other words, S satisfies Axioms $0 - 4$ of *Foundations.*

THEOREM 10. *Suppose that P is a point of a continuum K such that (1) K is locally compact at P and (2) K is aposyndetic at P. Then K together with the sum of all but a finite number of the complementary domains of K is connected im kleinen at P.*[14]

Indication of proof. Let L denote K together with the sum of all but a finite number of the complementary domains of K. Suppose that P is not an endpoint of S and that L is not connected im kleinen at P. There exists a connected domain U containing P such that (1) $\bar{U} \cdot K$ is compact, (2) the component of $U \cdot L$ which contains P contains no open subset of L containing P, and (3) the boundary β of $U \cdot K$ with respect to K contains no endpoint of S.[15] Let J denote a simple closed curve separating P from β,[16] let E denote the complementary domain of J that contains P, and let V denote the component of $E \cdot U$ that contains P. For each point X of F, the boundary of $V \cdot K$ with respect to K, there exists an open subset d_X of K which contains P and lies in K_X, a subcontinuum of K lying in $K - X$. Let G denote

[13] *Loc. cit.*[2]

[14] Cf. Theorem 36 on page 119 of *Foundations.*

[15] Part (3) of this statement is a direct consequence of Theorem 24 of my paper, "Certain consequences of the Jordan curve theorem," *American Journal of Mathematics*, vol. 63 (1941), pp. 531-544. This paper will be referred to as JCT.

[16] The existence of J follows from Theorem 2, part (3), of JCT and the argument for Theorem 3 on page 58 of my paper, "Concerning certain topologically flat spaces," *Transactions of the American Mathematical Society*, vol. 42 (1937), pp. 53-93.

the collection of continua K_X of K for all points X of F. Since F is closed and compact, there exists a finite subcollection $K_{X_1}, K_{X_2}, \cdots, K_{X_{\bar{n}}}$ of G such that $K - K_{X_1}, K - K_{X_2}, \cdots, K - K_{X_{\bar{n}}}$ covers F. Let D denote a connected domain lying in V and containing P such that $D \cdot K$ is a subset of Πd_{X_i}, $(i = 1, 2, \cdots, \bar{n})$. It follows from (2) above that there exists an infinite sequence $u_1, u_2, u_3, \cdots$ of mutually separate subsets of $U \cdot L$ no one of which contains P such that (a) for each integer n, $U \cdot L$ is the sum of two mutually separate sets Σu_i, $(1 \leqq i \leqq n)$, and $U \cdot L - \Sigma u_i$, $(1 \leqq i \leqq n)$, and (b) the limiting set of $u_1, u_2, u_3, \cdots$ contains P. Let D_1 denote a connected domain containing P which lies both in D and in a region of G_1 of Axiom 1 of *Foundations*. There exist an integer n_1 and an arc A_1B_1 which lies in D_1 such that A_1B_1 is irreducible from Σu_i, $(1 \leqq i \leqq n_1)$, to $U \cdot L - \Sigma u_i$, $(1 \leqq i \leqq n_1)$. Let D_2 denote a connected domain containing P which lies both in D_1 and in a region of G_2 of Axiom 1. There exists an integer n_2 such that $n_2 > n_1$ and Σu_i, $(1 \leqq i \leqq n_2)$, contains a point of D_2, and there exists in D_2 an arc irreducible from Σu_i, $(1 \leqq i \leqq n_2)$, to $U \cdot L - \Sigma u_i$, $(1 \leqq i \leqq n_2)$. This process may be continued and a sequence of arcs $A_1B_1, A_2B_2, A_3B_3, \cdots$ obtained such that (1) for each integer i, A_iB_i is a subset of D, (2) the sequential limiting set of $A_1B_1, A_2B_2, A_3B_3, \cdots$ is P, and (3) for each integer i, A_iB_i is irreducible between two mutually separate subsets of $U \cdot L$ whose sum is $U \cdot L$ such that A_i does not belong to the one that contains P. It follows from (3) that for each integer i, A_i and B_i belong respectively to distinct components h_i and k_i of $V \cdot K$ such that $\bar{h}_i \cdot \bar{k}_i = 0$ but for each integer i, $A_iB_i - (A_i + B_i)$ is a subset of $S - L$. It follows from Theorem 17 on page 167 of *Foundations* that there exist an infinite sequence of integers $m_1 < m_2 < m_3 \cdots$ and an infinite sequence of intervals $I_1, I_2, I_3, \cdots$ of J such that (a) if $i \neq j$, $\bar{h}_{m_i}$ contains no point of $\bar{h}_{m_j}$, (b) for each integer i, I_i contains $I_{i+1} + J \cdot \Sigma \bar{h}_{m_j}$, $(j \geqq i)$, but no point of $\bar{h}_{m_{i-1}}$ (when $i - 1$ is not zero), and (c) for each integer i, $A_{m_i}B_{m_i} - (A_{m_i} + B_{m_i})$ belongs to the same component, C, of $S - L$. The limiting set of $J \cdot \bar{h}_{m_1}, J \cdot \bar{h}_{m_2}, J \cdot \bar{h}_{m_3}, \cdots$ is non-vacuous and contains at most two points. Suppose that it contains only one point X. Since K contains a continuum (of G) containing all of the points $A_{m_1}, A_{m_2}, A_{m_3}, \cdots$ but not X, this is impossible. Hence the limiting set of $J \cdot \bar{h}_{m_1}, J \cdot \bar{h}_{m_2}, J \cdot \bar{h}_{m_3}, \cdots$ contains two points X and Y. Let M_X and M_Y denote elements of G containing $D \cdot K$ but *not* containing X and Y respectively. Let S_X and S_Y denote two segments of J containing X and Y respectively such that $\bar{S}_X \cdot \bar{S}_Y = \bar{S}_X \cdot M_X = \bar{S}_Y \cdot M_Y = 0$. Select two integers r and s from the set $m_1, m_2, m_3, \cdots$ such that (1) for two distinct values of i, $r < m_i < s$, and (2) for some integer i, $m_i < r$ and

$J \cdot \bar{h}_{m_i}$ is a subset of $S_X + S_Y$. Let O_rO_s denote an arc in C irreducible from A_rB_r to A_sB_s, and let P_r and P_s denote the first and last points in the order from O_r to O_s that O_rO_s has in common with J. Suppose that P_r belongs to S_X. There exist connected domains R, C_r, D_r, C_s, and D_s such that (1) R contains M_Y but $\bar{R} \cdot (\bar{S}_Y + O_rO_s) = 0$, (2) C_r, D_r, C_s, and D_s lie in V and contain h_r, k_r, h_s, and k_s respectively, (3) $\bar{C}_r$, $\bar{D}_r$, $\bar{C}_s$, and $\bar{D}_s$ are mutually exclusive, and (4) no component of $R \cdot \bar{V}$ contains a point of more than one of the sets $\bar{C}_r$, $\bar{D}_r$, $\bar{C}_s$, and $\bar{D}_s$. Let c_r, d_r, c_s, and d_s denote mutually exclusive connected domains such that the common part of each with J is a non-vacuous connected subset of S_Y and for each Z and j, ($Z = C$ or D and $j = r$ or s), $(\bar{R} + O_rO_s + \bar{D} + \bar{C}_r + \bar{D}_r + \bar{C}_s + \bar{D}_s - \bar{Z}_j) \cdot \bar{z}_j = 0$ but $Z_j \cdot z_j \neq 0$. There exists in $C_r + c_r + A_rB_r + D_r + d_r$ an arc $A'_rB'_r$ irreducible from $c_r \cdot S_Y$ to $d_r \cdot S_Y$. Except for its endpoints, $A'_rB'_r$ lies wholly in E and contains O_r. Let W denote an arc in $R + (A_rB_r - O_r)$ irreducible from the interval A'_rO_r of $A'_rB'_r$ to the interval $O_rB'_r$ of $A'_rB'_r$. Since W does not intersect S_Y, W contains two mutually exclusive arcs W_1 and W_2 irreducible from $A'_rB'_r$ to J which together contain the endpoints of W and abut on $A'_rB'_r$ from the same side. Also W_1 and the interval O_rP_r of O_rO_s abut on $A'_rB'_r$ from the same side, since $J - S_Y$ contains a point of both W_1 and O_rP_r but no point of $S_Y + A'_rB'_r$. Hence the complementary domain of the simple closed curve J_1 in $W + A'_rB'_r$ which does not contain Y contains $O_rO_s - O_r$ but no point of S_Y. Consequently O_rO_s contains no point of S_Y, and P_s belongs to S_X. Let $A'_sB'_s$ denote an arc in $C_s + c_s + A_sB_s + D_s + d_s$ irreducible from $c_s \cdot S_Y$ to $d_s \cdot S_Y$. Since $A'_sB'_s$ contains O_s, it is evident that W intersects both of the intervals A'_sO and $O_sB'_s$ of $A'_sB'_s$. Let W' denote an interval of W irreducible from A'_sO_s to $O_sB'_s$. For the same reasons that W intersected A'_sO_s and $O_sB'_s$, W' intersects A'_rO_r and $O_rB'_r$. Since W' contains no point of $A'_rB'_r$, this is a contradiction. So in case P is not an endpoint of S, L is connected im kleinen at P.

If P is an endpoint of S, it follows from Theorem 23 of JCT and the argument for Theorem 36 on page 119 of *Foundations*, that K and L are both connected im kleinen at P.

Examples. The reader should note that if the condition of local compactness at P is omitted from the hypothesis of Theorem 10, the resulting proposition is false. This is illustrated by the example following Theorem 3. Furthermore, if in Theorem 10 the condition that K be aposyndetic at P is replaced by the condition that K be semi-locally connected at P, the resulting proposition is false. This is easily seen to be the case for a continuum K lying in an Euclidean plane consisting of a straight line interval OP together

with the straight line intervals $OP_1, OP_2, OP_3, \cdots$ such that $P_1, P_2, P_3, \cdots$ converges to P and for each integer n, the angle POP_n is $1/n$. Although K is compact, semi-locally connected at P, and has only one complementary domain, K is not connected im kleinen at P.

Theorem 10 together with a theorem of the author[17] establishes the following two theorems. Theorem 12 is not only an extension of the classic Torhorst theorem to continua more general than continuous curves but also to spaces more general than spaces satisfying Axioms 0 — 5 of *Foundations* (which include the plane).

THEOREM 11. *Suppose that K is a locally compact continuum lying in the boundary of a connected domain D. Then in order that K be a continuous curve it is necessary and sufficient that K be a subset of a locally compact aposyndetic continuum M which contains no point of D.*

THEOREM 12. *Every component of the boundary of a complementary domain of a locally compact aposyndetic continuum is a locally compact continuous curve.*[18]

THE UNIVERSITY OF TEXAS.

---

[17] Theorem 1 of my paper, "Concerning the boundary of a complementary domain of a continuous curve," *Bulletin of the American Mathematical Society*, vol. 45 (1939), pp. 428-435.

[18] Cf. *Foundations*, p. 212; G. T. Whyburn, "Semi-locally connected sets," *loc. cit.*[3], p. 747; R. L. Wilder, "Sets which satisfy certain avoidability conditions," *loc. cit.*[3], p. 196; and R. L. Wilder, "Property S_n," *loc. cit.*[3], p. 832.

STRONGLY ARCWISE CONNECTED SPACES.*

By D. W. Hall and W. T. Puckett, Jr.

1. In a recent paper [1] the authors considered the problem of determining the class of all cyclic locally connected continua A such that for every arc-preserving transformation (continuity not assumed) $T(A) = B$, the set B is an arc or the transformation T is a homeomorphism. It was found that this class consists exactly of all cyclic continua A having the property that for every infinite subset M of A there exists an arc in A containing infinitely many points of M. Such spaces were called *strongly arcwise connected* and the trivial result that every strongly arcwise connected set is necessarily both compact and locally connected was pointed out.

An example of a cyclic locally connected continuum which fails to be strongly arcwise connected follows. Let L be the closed interval $(0, 1)$ and define $\{\alpha_i\}$ as a sequence of arcs having 0 as one endpoint but otherwise disjoint by pairs. Suppose further that $\delta(\alpha_i) \to 0$ and that for each i, α_i has the point $1/i$ as its other endpoint, while no point of L is an interior point of any α_i. Define

$$M = L + \Sigma\alpha_i.$$

Then M fails to be strongly arcwise connected. To see this we need only observe that if $\{x_i\}$ is a sequence of points in M converging to 0 and such that each x_i is an interior point of a different α_j then no arc in M contains infinitely many x_i. However, if $p \in M - 0$ and $x_i \to p$, it is easily seen that infinitely many x_i lie on some arc of M. Consequently, we say that M is strongly arcwise connected at every point except 0, but fails to be strongly arcwise connected at this point. In general, the locally connected continuum A will be said to be *strongly arcwise connected at the point* p provided that for any sequence of points $\{x_i\}$ converging to p some arc of A contains infinitely many x_i.

The purpose of this paper is to obtain two characterizations of strongly arcwise connected continua. One of these characterizations will show that a cyclic locally connected continuum A which fails to be strongly arcwise con-

* Received September 3, 1940; Presented to the American Mathematical Society, September 10, 1940.

[1] To appear in the *Bulletin of the American Mathematical Society.*

nected at a point p is essentially like the example above. From this characterization we shall deduce, among other things, that a cyclic locally connected continuum can fail to be strongly arcwise connected at an at most countable set of its points. That this is not true in general can be easily seen by considering a dendrite D which is the closure of its set of endpoints.

Our other characterization states that a locally connected continuum A is strongly arcwise connected if and only if for any infinite collection $[V]$ of open sets in A there exists an arc in A intersecting infinitely many of these sets.

2. In this section we obtain results useful in the proofs of our principal theorems.

THEOREM (2.1). *In the locally connected continuum A let U be a region such that $\bar{U}$ is a locally connected continuum and p a point of U. Then there exists a region $V(p)$ in U such that for any arc $\alpha = axb$ with $\alpha \cdot V \neq 0$ and $\alpha \cdot (A - U) = a + b$, the set $\bar{U}$ contains an arc apb.*[2]

Proof: The components of $\bar{U} - p$ intersecting $\bar{U} - U$ fall into two classes: (a) $X_1, X_2, \cdots, X_k$ such that p is an endpoint of each $\bar{X}_i$ and (b) $Y_1, Y_2, \cdots, Y_j$ such that p is not an endpoint of any $\bar{Y}_i$. Clearly, k and j are finite and either class may be vacuous. In each $\bar{X}_i \cdot U$ choose a region V_i containing p such that $F(V_i)$ is a single point. For each $\bar{Y}_i$ let K_i be the set of points separating some point of $\bar{U} - U$ from p in $\bar{Y}_i$. Since p does not cut $\bar{Y}_i$ and is not an endpoint of this set, p is not a limit point of K_i. Let W_i be a region in $\bar{Y}_i$ containing p, of diameter $\leqq d/2$, where $d = \min(p, K_i)$. Set $V = V_1 + \cdots + V_k + W_1 + \cdots + W_j +$ all components of $\bar{U} - p$ not intersecting $\bar{U} - U$. Every arc α such that $\alpha \cdot V \neq 0$ and $\alpha \cdot (\bar{U} - U) = a + b$ either passes through p or, by the choice of d, has a non-degenerate intersection with the true cyclic element C_i of $\bar{Y}_i$ which contains p, where $\bar{Y}_i$ contains $a + b$. An application of the Three-point Theorem gives the desired result.

THEOREM (2.2). *Let p be any point of the locally connected continuum A and define $A^* = A - p$. If x_1, x_2 and y are three points of A^* such that* (a) *no point separates y and $x_1 + x_2$ in A^*, and* (b) *x_i does not separate y from x_j $(i \neq j)$ in A^*, then A^* contains an arc x_1x_2 having y as an interior point.*

Proof: Suppose first that x_1, x_2 and y lie in the same nodule[3] E of A^*.

[2] The proof given for (2.1) is due to the referee, and is somewhat shorter than the one found by the authors.

[3] See G. T. Whyburn, "On the structure of connected and connected im kleinen point sets," *Transactions of the American Mathematical Society*, vol. 32 (1930), pp. 926-943, in particular pp. 929 and 933. We shall refer to this paper hereafter as N.

Now [4] E is closed in A^*, connected and locally connected. Therefore, since E is locally compact and has no cut point of itself, it follows [5] that E is cyclic. Hence by the Three-point Theorem there exists an arc x_1yx_2 in E.

If x_1, x_2 and y do not lie in the same nodule of A^*, then there exists a point q separating two of these points in A^*. By (b) q is neither x_1 nor x_2. If q is y, then any arc x_1x_2 in A^* has y in its interior and the proof is complete. Therefore, assume $q \neq y$ and let C be the simple nodular chain [6] joining x_1 and x_2 in A^*. Since any component of $A^* - C$ has a single point as its boundary,[7] it follows that y is a point of C. Now [8] C is connected, locally connected, and by the above reasoning y is not a cut point of C. Thus there exists a nodule [9] E of C containing y. Let x_1x_2 be an arc in C, and define z_1 and z_2 to be the first and last points, respectively, of x_1x_2 in E. By the first paragraph of the proof there will exist an arc z_1yz_2 in E. Replacing the subarc z_1z_2 of x_1x_2 by z_1yz_2 will give the required arc x_1yx_2 in A^*.

THEOREM (2.3). *Let there exist in the locally connected continuum A a sequence of points $\{x_n\}$ converging to p and a sequence of points $\{p_n\}$ such that for each n*

$$A - (p + p_n) = K_n + A_n + M_n,$$

where the component K_n contains $\sum_{i=1}^{n} x_i$ and the component A_n contains $\sum_{i=n+1}^{\infty} x_i$. If for each n no point separates x_n and $p_{n-1} + p_n$ in $A - p$, then $\{x_n\}$ lies on an arc in A.

Proof: For each n the points p_{n-1}, p_n and x_n satisfy the conditions of (2.2) since $p_{n-1} + x_n$ lies in K_n and $p_n + x_n$ lies in A_{n-1}. Hence for each $n > 1$ there exists an arc $\alpha_n = p_{n-1}x_np_n$ in $A - p$. Moreover, α_n is contained in $\bar{K}_n \cdot \overline{A_{n-1}}$, since $F(K_n)$ lies in $p_n + p$, $F(A_{n-1})$ lies in $p_{n-1} + p$, and p is not a point of α_n. Let $\alpha_1 = x_1p_1$ be an arc in $\bar{K}_1 - p$, then $\alpha_j \cdot \alpha_k = 0$ for $j \leq k - 2$ and $\alpha_{k-1} \cdot \alpha_k = p_{k-1}$. Now each K_i contains at most a finite number of p_i and, consequently, can contain no limit point of Σp_i. Therefore, the interior of each α_i lies in a single component of $A - (\overline{p + \Sigma p_i}) = A' + Q_1 + Q_2 + \cdots$, where α_i lies in $\bar{Q}_i$. Now $\delta(Q_i)$ converges to 0, since $\{Q_i\}$ is a null sequence. Hence $\alpha = \overline{\Sigma \alpha_i}$ is an arc containing all the x_i.

---

[4] See N (3.5).

[5] See G. T. Whyburn, "On the cyclic connectivity theorem," *Bulletin of the American Mathematical Society*, vol. 37 (1931), pp. 429-433.

[6] See N (5.3) and (4.1).

[7] See N (6.4).

[8] See N (5.4).

[9] See N (4.1) and (3.6).

THEOREM (2.4). *In the locally connected continuum A let there exist separations*

$$A - (p + p_i) = K_i + A_i + M_i; \qquad (i = 1, 2, 3),$$

where K_i and A_i are components with $p_i \in F(K_i) \cdot F(A_i)$, and suppose there exists a point $y \in \prod_{i=1}^{3} A_i$, and points $x_i \in A_{i-1}K_i$ $(i = 2, 3)$. (a) *If $K_1 \subset K_2$ and q separates x_2 and $p_1 + p_2$ in $A - p$, then*

$$A - (p + q) = K'_2 + A'_2 + M'_2,$$

where K'_2 and A'_2 are components with $x_2 \in K'_2$, $A'_2 \supset A_2$, and $K'_2K_1 = 0$. (b) *If $K_1 \subset K_2K_3$, then $K_2 \subset K_3$.*

Proof: To prove (a) we note that $K_2 + p_2$ is connected and contains both x_2 and p_2. Therefore, K_2 contains q, and thus $q \neq y$. Let A'_2 be the component of $A - (p + q)$ containing y. Since A_2 intersects A'_2 in y and does not contain q, we have $A'_2 \supset A_2$. Moreover, $p_2 \in A'_2$, hence x_2 does not lie in this set. Let K'_2 be the component of $A - (p + q)$ containing x_2. Suppose $K'_2K_1 \neq 0$, then since $x_2 \in K'_2A_1$, we must have $p_1 \in K'_2$. This is impossible hence the proof of (a) is complete.

Considering (b) we have at once that A_2 intersects both K_3 and A_3, hence $p_3 \in A_2$. Thus p_3 is not in K_2; hence, since $K_2K_3 \neq 0$, we have $K_2 \subset K_3$.

THEOREM (2.5). *If the sequence of points $\{x_i\}$ converges to a point p in the locally connected continuum A and if no point separates any two x_i in $A - p$, then infinitely many x_i lie on an arc in A.*

Proof: Evidently the x_i lie in a single cyclic element E_1 of A. Moreover, p is a point of E_1 and under our hypotheses no point separates any two x_i in $E_1 - p$. Let R_1 be a region in E_1 containing p and such that $\delta(R_1) < 1$, $\bar{R}_1$ is a locally connected continuum, and x_1 is not a point of $\bar{R}_1$. Clearly we may assume that R_1 contains all the x_i for $i > 1$. Finally, we may assume that all x_i for $i > 1$ lie in a single component M of $\bar{R}_1 - p$, since there can be only a finite number of components. Moreover, $\bar{M}$ is a locally connected continuum.

Let us suppose that infinitely many x_i do not lie on an arc in A. Then we shall show that infinitely many x_i lie in a single cyclic element E_2 of $\bar{M}$ (consequently, of $\bar{R}_1$) and have the property that no point separates any two of them in $E_2 - p$. Suppose a point p_1 does separate two x_i $(i > 1)$ in $\bar{M} - p$. Then $\bar{M} - (p + p_1) = K_1 + M_1 + L_1$ is a separation for which it may be assumed that the component K_1 contains x_2 and that the component M_1 contains all x_i for $i > 2$. Moreover, both K_1 and M_1 intersect $F(R_1)$, since no

point separates two x_i in $E_1 - p$. Likewise, no point separates two x_i $(i > 2)$ in $\bar{M} - p$ or there exists a point p_2 and a separation $\bar{M} - (p + p_2) = K_2 + M_2 + L_2$, where the component K_2 contains x_3, the component M_2 contains all x_i for $i > 3$, and both K_2 and M_2 intersect $F(R_1)$. Now K_1 must lie in some component of $\bar{M} - (p + p_2)$, since p_2 is in M_1. If K_1 is not contained in K_2 then $K_1K_2 = 0$ and, consequently, K_1 and K_2 are components of $\bar{M} - (p + p_1 + p_2)$ intersecting $F(R_1)$. If $K_1 \subset K_2$ and there exists a point q_2 which separates x_3 and $p_1 + p_2$ in $\bar{M} - p$, then by (2.4a) we may find in $\bar{M} - (p + q_2)$ a component M'_2 containing M_2 and a component K'_2 containing x_3 such that $K'_2 \cdot K_1 = 0$. Thus we may assume that either $K_2K_1 = 0$, or $K_1 \subset K_2$ and no point separates x_3 and $p_1 + p_2$ in $\bar{M} - p$.

Generally, either no point separates two x_i $(i > n - 1)$ in $\bar{M} - p$ or there exists a separation $\bar{M} - (p + p_n) = K_n + M_n + L_n$, where the component K_n contains x_{n+1}, the component M_n contains all x_i $(i > n + 1)$, and both K_n and M_n intersect $F(R_1)$. Moreover, after (2.4a) it may be assumed that either $K_iK_n = 0$ for all $i < n$ or for some $k < n$ we have $K_k \subset K_n$ and no point separates x_{n+1} and $p_k + p_n$ in $\bar{M} - p$. In any case ΣK_i cannot contain a limit point of Σp_i, since M_n clearly contains all p_i for $i > n$. Now suppose infinitely many of the K_i, say $\{K_{a_i}\}$, have the property $K_{a_i} \cdot K_{a_j} = 0$ for $i \neq j$. Then $\bar{M} - (\overline{p + \Sigma p_{a_i}})$ has infinitely many components K_{a_i} intersecting $F(R_1)$ and having p in their limit inferior, which is impossible. We can thus assume that for each $n > 1$ there exists a $k < n$ such that $K_k \subset K_n$ and no point separates x_{n+1} and $p_k + p_n$ in $\bar{M} - p$. Thus $K_1 \subset K_2$ and after (2.4b) $K_1 \subset K_2 \subset K_3$. Thus we have generally by (2.4b) $K_1 \subset K_2 \subset \cdots \subset K_n$. But this must stop for some finite n for otherwise by (2.3) infinitely many x_i would lie on an arc. Thus in any case M contains infinitely many x_i such that no point separates any two of them in $\bar{M} - p$. Therefore, $\bar{M}$ contains a cyclic element E_2 with these same properties.

The supposition that no arc in A contains infinitely many x_i and the above reasoning give us for each n a cyclic element E_n of $\bar{R}_{n-1}$ where E_n and R_{n-1} have the following properties: (i) R_{n-1} is a region about p in E_{n-1} and $\bar{R}_{n-1}$ is a locally connected continuum, (ii) $\delta(R_{n-1}) < 1/(n-1)$, and (iii) E_n contains all x_i $(i \geq n)$ and no point separates any two of these points in $E_n - p$.

By (2.2) there exists an arc $\alpha_0 = x_2x_1x_3$ in $E_1 - p$. Since p is not a point of α_0, there exists a first x_i, say x_{i_1} which is not on α_0. Moreover, by (2.2) there exists an arc $\alpha_1 = x_2x_{i_1}x_3$ in $E_2 - p$. The set $\alpha_0 + \alpha_1$ contains a simple closed curve J_1 through $x_1 + x_{i_1}$ which intersects E_2 in more than one point. Thus $A_1 = J_1 + E_2$ contains all the x_i and no point separates any two

of them in $A_1 - p$. Let E_{i_2} be the first E_i not intersecting J_1, then $x_{i_2} \epsilon E_{i_2}$ and by (2.2) there exists an arc $\alpha_2 = x_1 x_{i_2} x_{i_1}$ in $A_1 - p$. Hence $\alpha_2 E_2$ contains an arc $\beta_2 = y_1 x_{i_2} z_1$ spanning J_1. Define $J_2 = J_1 + \beta_2$, then $x_{i_1} + x_{i_2} \subset J_2 E_{i_1}$. Let E_{i_3} be the first E_i not intersecting J_2 and by (2.2) construct an arc $\alpha_3 = x_{i_1} x_{i_3} x_{i_2}$ in $E_{i_1} - p$. Since $J_1 E_{i_1} = 0$, the arc α_3 must contain a subarc $\beta_3 = y_2 x_{i_3} z_2$ spanning β_2. Define $J_3 = J_2 + \beta_3 = J_1 + \beta_2 + \beta_3$, then $x_{i_2} + x_{i_3} \subset E_{i_2} \cdot J_3$.

Generally for each $n > 2$ let E_{i_n} be the first E_i not intersecting J_{n-1} and by (2.2) construct an arc $\alpha_n = x_{i_{n-2}} x_{i_n} x_{i_{n-1}}$ in $E_{i_{n-2}} - p$. Since $J_{n-2} E_{i_{n-2}} = 0$, the arc α_n must contain a subarc $\beta_n = y_{n-1} x_{i_n} z_{n-1}$ spanning β_{n-1}. Define $J_n = J_{n-1} + \beta_n = J_1 + \sum_{i=2}^{n} \beta_i$, then $x_1 + \sum_{j=1}^{n} x_{i_j} \subset J_n$ and $x_{i_{n-1}} + x_{i_n} \subset E_{i_{n-1}}$. Since $E_i \to p$, it follows that $\beta_i \to p$ and, consequently, $J = p + \sum_{n=1}^{\infty} J_n$ is a locally connected continuum. In J_1 define the subarcs $\gamma_1 = x_1 y_1 z_1$ and $\mu_1 = x_1 y_1 z_1$, while in each β_i $(i = 2, 3, \cdots)$ define the subarcs $\gamma_i = y_{i-1} z_i y_i$ and $\mu_i = z_{i-1} y_i z_i$, renaming the y's and z's if necessary. Let $M = p + \sum_{n=1}^{\infty} (\gamma_{2n-1} + \mu_{2n})$ and $N = p + \sum_{n=1}^{\infty} (\gamma_{2n} + \mu_{2n-1})$, then M and N are arcs such that $M + N = J$. But J contains infinitely many x_i. Thus either M or N contains infinitely many x_i, and the proof of (2.5) is complete.

3. We now come to the proofs of our principal results.

THEOREM (3.1). *In order that the locally connected continuum A be strongly arcwise connected it is necessary and sufficient that for any infinite collection of open sets $[V]$ in A there exists an arc in A intersecting infinitely many of these sets.*

Proof: The necessity of the condition follows at once from the definition of strong arcwise connectivity. Let $\{x_i\}$ be a sequence of points in A converging to the point x, then in order to complete the proof of the theorem we must show that there exists an arc in A containing infinitely many of the points x_i. For each i let U_i be a region containing x_i such that $\{U_i\}$ is a null sequence, x is not a point of ΣU_i, and $\bar{U}_i \bar{U}_j = 0$ $(i \neq j)$. By (2.1) each U_i contain a region V_i about p such that for any arc $\beta = ayb$, where y is a point of V_i and $\beta \cdot (A - U_i) = a + b$, there exists an arc ax_ib in $\bar{U}_i$. Under our hypothesis there exists an arc α in A intersecting infinitely many of the sets V_i. By construction $\lim V_i = x$. We may thus assume that x is an end-point of α, and that α intersects all the V_i. Denote the other endpoint of α by y and let

U_{i_1} be the first U_i not containing y. If x_{i_1} is not a point of α, then by (2. 1), $\bar{U}_{i_1}$ contains an arc $p_1x_{i_1}q_1$ spanning α. In this case $\alpha - (p_1 + q_1)$ has the components X_1 containing x, Y_1 containing y, and Z_1. Moreover, X_1 must intersect infinitely many V_i since x is not a point of $Y_1 + Z_1$. Hence $\alpha_1 = X_1 + p_1x_{i_1}q_1 + Y_1$ ($\alpha_1 = \alpha$ if x_{i_1} is a point of α) is an arc through x_{i_1} intersecting infinitely many V_i $(i > i_1)$. Now there exists a V_{i_2} $(i_2 > i_1)$ such that $\alpha_1 \cdot V_{i_2} \neq 0$ and the subarc yx_{i_1} of α_1 lies in a single component of $\alpha_1 - \alpha_1\bar{U}_{i_2}$. Thus in case x_{i_2} is not a point of α_1 there exists in $\bar{U}_{i_2}$ an arc $p_2x_{i_2}q_2$ spanning α_1. Then $\alpha_1 - (p_2 + q_2)$ has components X_2 containing x and points from infinitely many V_i $(i > i_2)$, Y_2 containing yx_{i_1}, and Z_2. Hence $\alpha_2 = X_2 + p_2x_{i_2}q_2 + Y_2$ ($\alpha_2 = \alpha_1$ if x_{i_2} is a point of α_1) is an arc containing $x_{i_1} + x_{i_2}$ and intersecting infinitely many V_i $(i > i_2)$. In the same way we obtain for every k an arc $\alpha_k = X_k + p_kx_{i_k}q_k + Y_k$ ($\alpha_k = \alpha_{k-1}$ if x_{i_k} is a point of α_{k-1}) containing $\sum_{j=1}^{k} x_{i_j}$ and intersecting infinitely many V_i $(i > i_k)$. Now the continuum $\beta = \lim \alpha_k$ contains x, y and infinitely many x_i. Moreover any point z of β lies on α_k for almost all k, since any modification of α_k takes place in $\sum_{j=i_k+1}^{\infty} \bar{U}_j$ and $\lim \bar{U}_i = x$. Therefore, β is an arc containing infinitely many x_i, and the proof is complete.

THEOREM (3. 2). *In order that the locally connected continuum A fail to be strongly arcwise connected at the point p it is necessary and sufficient that there exist a closed subset N of A containing p and a separation $A - N = M + \sum_{i=1}^{\infty} K_i$, where the K_i are distinct components of $A - N$ such that $K_i \rightarrow p$ and for each i, $F(K_i) \subset p + p_i \subset N$.*

Proof. For each i let x_i be a point of K_i, then the x_i converge to p. The sufficiency part of the theorem follows immediately, since not more than four of the x_i can lie on an arc in A.

Now assume that A is not strongly arcwise connected at p. Then there exists a sequence of points $x_i \rightarrow p$ such that no arc in A contains infinitely many of the x_i. We may assume that all x_i lie in a single component of $A - p$, for otherwise put $N = p$. By (2. 5) there exists a point p'_1 separating some x_k from infinitely many x_i in $A - p$, hence we may assume that p'_1 separates x_1 from all x_i $(i > 1)$ in this set. Moreover, we lose no generality in assuming that all x_i $(i > 1)$ lie in a single component A'_1 of $A - (p + p'_1)$. Then we have the separation $A - (p + p'_1) = K'_1 + A'_1 + M'_1$, where the component K'_1 contains x_1. Again by (2. 5) there exists a point p'_2 and a separation $A - (p + p'_2) = K'_2 + A'_2 + M'_2$, where we may assume the component K'_2

contains x_2 and the component A'_2 contains all x_i for $i > 2$. We observe that p'_2 is a point of A'_1. In general (2.5) gives a point p'_n and a separation $A - (p + p'_n) = K'_n + A'_n + M'_n$, which on renaming the x_i for $i > n - 1$ has the following properties:

(i) *For each n, the component K'_n contains x_n, the component A'_n contains all x_i for $i > n$, and $p'_n \epsilon A'_{n-1}$.*

(ii) *For each n, $p'_n \epsilon F(K'_n) \cdot F(A'_n)$ and $F(K'_n) + F(A'_n) = p + p'_n$.*

(iii) *Define $P' = \sum_{i=1}^{\infty} p'_i$, then $A'_n \cdot P' \supset \sum_{i=n+1}^{\infty} p'_i$ and $K'_n \cdot P' \subset \sum_{i=1}^{n-1} p'_i$.*

(iv) *If $j < k$, then K'_j lies in a single component of $A - (p + p'_k)$. Thus either $K'_j \cdot K'_k = 0$ or $K'_j \subset K'_k$.* This is true since $p'_k \epsilon A'_j$.

(v) *If $j < k$ and $K'_j \subset K'_k$, then it may be assumed that no point separates x_k and $p'_j + p'_k$ in $A - p$.*

If this were not true, then by (2.4) we could find a separation $A - (p + p''_k) = K''_k + A''_k + M''_k$ having properties (i)-(iii) and $K'_j \cdot K''_k = 0$. Thus by renaming p''_k, K''_k, A''_k, and M''_k we obtain the desired result.

(vi) *There exists a subsequence $\{K_i\}$ of $\{K'_i\}$ such that $K_j K_k = 0$ for $j \neq k$.*

We show first that there exist two sets K'_j and K'_k such that $K'_j \cdot K'_k = 0$. Otherwise it follows from (iv), (2.4b), and (v) that $K'_1 \subset K'_2 \subset \cdots \subset K'_n \subset \cdots$, where for each n no point separates x_n and $p'_{n-1} + p'_n$ in $A - p$. Thus by (2.3) all the x_i lie on an arc, which is impossible.

Now let $K_1, K_2, \cdots, K_n$ be an ordered subcollection of $K'_1, K'_2, \cdots, K'_m$ $(m \geq n)$ such that $K_j K_k = 0$ for $j, k = 1, 2, \cdots, n$ and $j \neq k$. The proof of (vi) will be complete if we can find among the K'_i $(i > m)$ a set, which we denote by K_{n+1}, such that $K_j K_k = 0$ for $j, k = 1, 2, \cdots, n + 1$ and $j \neq k$. To this end let $y_1, y_2, \cdots, y_n$ and $p_1, p_2, \cdots, p_n$ be the subcollections of $\{x_i\}$ and $\{p_i\}$ corresponding to $K_1, K_2, \cdots, K_n$. If no $K_i \subset K'_{m+1}$ we may by (iv) put $K_{n+1} = K'_{m+1}$ and the proof is complete.

In case K'_{m+1} contains a K_i, we consider K'_{m+2}. If K'_{m+2} contains no K_i, put $K_{n+1} = K'_{m+2}$ and again the proof is complete. Thus either the proof will be complete by the time we reach K'_{m+n+2} or there will exist a K_i, say K_k, such that $K_k \subset K'_{m+i} K'_{m+j}$ $(i < j < n + 2)$. Now by (2.4b) and (v) $K_k \subset K'_{m+i} \subset K'_{m+j}$, while no point separates x_{m+i} and $p_k + p'_{m+i}$ in $A - p$ and no point separates x_{m+j} and $p'_{m+i} + p'_{m+j}$ in this set. Let α be a positive

integer; then after at most αn steps the proof is complete or we obtain $K_k \subset K'_{m+i_1} \subset K'_{m+i_2} \subset \cdots \subset K'_{m+i_\alpha}$ $(i_1 < i_2 < \cdots < i_\alpha < \alpha n)$ where for each $j > 1$ no point separates x_{m+i_j} and $p'_{m+i_{j-1}} + p'_{m+i_j}$ in $A - p$. But since no arc contains infinitely many x_i, it follows from (2.3) that this monotone increasing sequence of sets can contain but a finite number of distinct terms. Hence after a finite number of steps we must obtain a K'_i, call it K_{n+1}, with the property $K_i \cdot K_j = 0$ $(i, j = 1, 2, \cdots, n+1$ and $i \neq j)$.

Take $\{K_i\}$ as the subsequence of $\{K'_i\}$ given by (vi), $\{p_i\}$ as the corresponding subsequence of $\{p_i\}$ and $N = \overline{p + \Sigma p_i}$. Now $K_i \cdot K_j = 0$ $(i \neq j)$ and $K_i \cdot \overline{\Sigma p_j} = 0$, since K_i is open and $F(K_j) \supset p_j$. Hence $A - N = M + \Sigma K_i$ is the desired decomposition.

In the following corollaries the set A is a locally connected continuum.

Corollary (3.21). *If no two points separate A then A is strongly arcwise connected.*

Corollary (3.22). *If A is cyclic and fails to be strongly arcwise connected at a point p then* (a) *p is a local separating point of A,* (b) *p is not of finite order in A,* (c) *p is an im kleinen cycle point*[10] *of A.*

Corollary (3.23). *If A is cyclic then the set of points where A fails to be strongly arcwise connected is at most countable.*[11]

Brown University.
University of California at Los Angeles.

[10] See G. T. Whyburn, *Mathematische Annalen*, vol. 102 (1929), p. 317, Theorem 6.
[11] See G. T. Whyburn, *loc. cit.*[10], Theorem 5.

GREEN'S LEMMA AND RELATED RESULTS.* [1]

By William T. Reid.

1. Introduction. The proof of Green's lemma for the case of a region R which is the interior of a general simply closed rectifiable curve J involves certain considerations, largely topological in character, which are not necessary when the boundary of the region is restricted to be relatively simple in character. Proofs of Green's lemma for such a general region have been given by Gross [5],[2] Van Vleck [14], Pollard [13] and Bray [1]. The demonstrations of Pollard and Bray involve, among other things, the chain-theorems of de la Vallée Poussin. Bray imposes continuity and differentiability conditions on the functions involved not only interior to and on the simply closed curve J, but also in a neighborhood of this point set; it is to be noted, however, that he obtains each of the individual "partial integration" formulas of Green's lemma under a weaker condition than the rectifiability of J. The demonstrations of Gross and Van Vleck are extremely tedious in detail.

There is a corresponding general form of Cauchy's theorem for a function $f(z)$ which is holomorphic on the interior R of a simply closed rectifiable curve J, and merely continuous on $R + J$. The proof of this result given by Pollard [13] involves de la Vallée Poussin's chain theorems. The extremely brief demonstration given by Heilbronn [6] utilizes such powerful tools as the Riemann mapping theorem, a theorem on polynomial approximation due to Fejér, as well as results for the Stieltjes integral. For an intermediate case in which $f(z)$ is assumed not merely to be continuous, but to be regular, on the closed set $R + J$ various proofs have been given (see, for example, Goursat-Hedrick [3], pp. 66-70; E. H. Moore [10], and Kamke [9]).

The purpose of the present paper is to give a proof of Green's lemma for the general case of a region R which is the interior of a simply closed rectifiable curve J, and where the continuity and differentiability conditions imposed on the functions involved are merely on $R + J$ and R, respectively. The author believes that the details of proof are simpler than those involved in the previous demonstrations. In particular, aside from the Jordan curve theorem itself, the chief topological result used is that stated in Lemma I below. Analytically, Ascoli's theorem and a very special case of the general

* Received October 26, 1940.

[1] Presented to the American Mathematical Society, April 12, 1941.

[2] Numbers in square brackets refer to the bibliography at the end of this paper.

convergence theorem of Hobson are employed. Theorems 1 and 2 are concerned with the individual partial integration formulas occurring in Green's lemma. The general form of Cauchy's theorem for a function $f(z)$ which is holomorphic on R, and merely continuous on $R+J$, is presented in Theorem 3. Finally, Theorem 4 gives a result which has been announced by Hestenes [7] [3] under additional regularity assumptions on the boundary curve J.

2. Principal results. The main result of this paper is the following theorem.

THEOREM 1. *Suppose that R is the interior of a simply closed rectifiable plane curve J, and that $M(x,y)$ is such that*: (i) *it is continuous on $R+J$*; (ii) *if $a\leqq x\leqq b$, $c\leqq y\leqq d$ is any rectangle in R, $M(x,y_0)$ is absolutely continuous in x on $a\leqq x\leqq b$ for y_0 almost everywhere on $c\leqq y\leqq d$*; (iii) *$\partial M/\partial x$ is summable [Lebesgue integrable] on R.*[4] *Then*

$$\int\int_R \frac{\partial M}{\partial x}\,dxdy=\int_J M(x,y)\,dy, \tag{1}$$

the line integral being taken around J in the positive (counter-clockwise) direction.

Under the hypotheses of the theorem we clearly have

$$\int\int_{R'} \frac{\partial M}{\partial x}\,dxdy=\int_{J'} M(x,y)\,dy \tag{2}$$

if R' and J' denote the interior and boundary, respectively, of a rectangle $a\leqq x\leqq b$, $c\leqq y\leqq d$ lying in R. It also follows by a relatively simple argument that (2) is valid for R' the interior of a simply closed polygon J' lying in R, each of whose component straight line segments is parallel to either the x- or y-axis. This result may be established by drawing the finite number of lines, parallel to the coördinate axes, containing the separate sides of the simply closed polygon J', and integrating around each rectangle so defined, whose interior belongs to R', in a counter-clockwise direction. It may

[3] [Added in proof, May 23, 1941.] Hestenes never published this announced result. In a revised form of his original manuscript, to appear soon in the *Duke Mathematical Journal* under the title "An analogue of Green's theorem in the calculus of variations," Hestenes gives a proof of his previously announced result under conditions of the same generality as those of Theorem 4 of the present paper. The forthcoming paper of Hestenes also contains a proof of Green's lemma for the general case described above.

[4] It is a consequence of (i) and (ii) that $\partial M/\partial x$ exists almost everywhere and is measurable on R. See, for example, Carathéodory [2], p. 642.

also be proved by an inductive argument; of these two methods the author prefers the latter.

Now consider a general simply closed rectifiable curve J, and suppose that $O: (x_0, y_0)$ is a point of the interior R of J. Then there exists a segment AOB of the line $y = y_0$ which contains O, has its end-points A and B on J and which, except for these end-points, lies in R. Let BCA and $BC'A$ denote the two simple arcs of J having end-points at B and A. If the conclusion of Theorem 1 holds for each of the simply closed curves $AOBCA$, $AOBC'A$, then clearly this theorem is also valid for the original curve J. Hence without loss of generality we may restrict attention to a simply closed rectifiable curve of the type $AOBCA$, and for simplicity we shall do so. In particular, we have the following result whose proof will be deferred to the following section.

LEMMA I. *Suppose J is a simply closed rectifiable plane curve consisting of a straight line segment AOB of the line $y = y_0$, and a rectifiable simple arc $\Gamma: BCA$. Corresponding to a given $\epsilon > 0$ there exists a simple polygonal arc $\Gamma': B'C'A'$ such that*: (a) *B' and A' are interior to OB and AO, respectively*; (b) *Γ' lies, except for B' and A', in the interior R of J*; (c) *each straight line segment of Γ' is parallel to either the x- or y-axis*; (d) *the length of Γ' does not exceed five times the length of Γ*; (e) *each point of Γ' lies in the ϵ-neighborhood of Γ.*

Consider a simply closed curve $J: AOBCA$ of the sort to which Lemma I applies. We shall suppose that the letters A, B are so assigned that the order AOB is "counter-clockwise" along J. If $O: (x_0, y_0)$, clearly this notion can be stated entirely in terms of whether points (x_0, y) of the ray $x = x_0$, $y > y_0$ for y sufficiently near y_0 lie interior or exterior to J. Corresponding to $\epsilon = 1$, let $\Gamma_1: B_1C_1A_1$ denote a simple polygonal arc satisfying the conditions of Lemma I. We shall define a sequence of arcs $\Gamma_n: B_nC_nA_n$, $(n = 1, 2, \cdots)$, as follows. If arcs $\Gamma_1, \cdots, \Gamma_k$, $(k \geqq 1)$, have been defined, let J_j denote the simply closed polygon $A_jOB_jC_jA_j$, $(j = 1, \cdots, k)$, and let R_j be the interior of J_j; moreover, let d_k denote the distance of the arc $\Gamma: BCA$ from the point set $R_k + J_k$. Then Γ_{k+1} is chosen as an arc $B_{k+1}C_{k+1}A_{k+1}$ satisfying the conditions of Lemma I for $\epsilon = \operatorname{Min}\{d_k, 1/(k+1)\}$. Clearly A_{k+1} is interior to the arc AA_k and B_{k+1} is interior to B_kB; moreover, R_k and the interior points of Γ_k belong to the interior R_{k+1} of the simply closed polygon J_{k+1}: $A_{k+1}OB_{k+1}C_{k+1}A_{k+1}$. If p is an arbitrary point of R, then there exists an arc prO which lies, except for O, in R. If N is an integer such that $1/N$ does not exceed the distance between the disjoint arcs BCA and prO, then clearly p is a point of R_n for $n > N$. Consequently, $\Sigma_n R_n = R$; moreover, it is seen that A_n and B_n approach A and B, respectively, as $n \to \infty$.

Now Γ_n admits the analytic representation $x=\phi_n(s)$, $y=\psi_n(s)$, $(0\leqq s\leqq L_n)$, where s is the arc length along this polygon measured from B, and L_n is the length of Γ_n. Since $L_n\leqq 5L$, where L is the length of BCA, if one writes $s=tL_n$, $x_n(t)=\phi_n(tL_n)$, $y_n(t)=\psi_n(tL_n)$, then the arcs Γ_n admit parametrizations:

$$\Gamma_n: x=x_n(t),\ y=y_n(t), \qquad (0\leqq t\leqq 1;\ n=1,2,\cdots); \tag{3}$$

furthermore, $x_n(t)$, $y_n(t)$ satisfy a Lipschitz condition with uniform Lipschitz constant $5L$. Hence this set of functions is equi-continuous on $0\leqq t\leqq 1$, and by Ascoli's theorem (see, for example, [8], p. 168) there exist functions $x_0(t), y_0(t)$ and a sub-sequence $\{x_{n_k}(t), y_{n_k}(t)\}$ such that $\lim_{k\to\infty} x_{n_k}(t)=x_0(t)$, $\lim_{k\to\infty} y_{n_k}(t)=y_0(t)$ uniformly on $0\leqq t\leqq 1$. Clearly $x_0(t), y_0(t)$ are continuous and satisfy the Lipschitz condition with constant $5L$ on this interval. For convenience of notation, we shall assume that the original sequence of arcs Γ_n is so chosen that the coördinate functions of (3) satisfy these convergence properties.

Now consider the curve Γ_0 defined by

$$\Gamma_0: x=x_0(t),\ y=y_0(t), \qquad (0\leqq t\leqq 1). \tag{4}$$

Clearly this is a path curve with initial point $(x_0(0), y_0(0))=B$ and terminal point $(x_0(1), y_0(1))=A$; moreover, each point of Γ_0 lies on the arc Γ. Since Γ is disconnected by the omission of one of its interior points, and the continuity of $x_0(t), y_0(t)$ implies that the locus of Γ_0 is a connected set of points, it then follows that each point of Γ lies on the path curve Γ_0. We shall denote by J_0 the closed path curve consisting of the line segment AOB and Γ_0.

Let us now return to the relation (1). Because of the special character of J_n, it follows that

$$\iint_{R_n} \frac{\partial M}{\partial x}\,dxdy=\int_{J_n} M(x,y)\,dy, \qquad (n=1,2,\cdots). \tag{5}$$

Since each R_n is an open set, R_n is a subset of R_{n+1}, and $\Sigma_n R_n=R$, it follows that the measure of $R-R_n$ tends to zero as $n\to\infty$. As $\partial M/\partial x$ is supposed summable on R, it then follows that the limit of the first member of (5) as $n\to\infty$ is the first member of (1). Then

$$\begin{aligned}\int_{J_n} M(x,y)\,dy-\int_{J_0} M(x,y)\,dy\\ &=\int_0^1 M(x_n(t),y_n(t))y'_n(t)\,dt-\int_0^1 M(x_0(t),y_0(t))y'_0(t)\,dt\\ &=\int_0^1 [M(x_n,y_n)-M(x_0,y_0)]y'_n\,dt+\int_0^1 M(x_0,y_0)[y'_n-y'_0]\,dt. \qquad (6)\end{aligned}$$

Now the first of the integrals of (6) approaches zero as $n \to \infty$ since $|y'_n(t)| \leqq 5L$ and $M(x_n(t), y_n(t)) - M(x_0(t), y_0(t)) \to 0$ uniformly on $0 \leqq t \leqq 1$. Since $y_n(t) \to y_0(t)$ uniformly on this interval, the convergence of the second integral to zero is a consequence of a general convergence theorem of Hobson ([8], p. 422). For a proof of a special instance of this theorem, which suffices for the present case, the reader is referred to Graves [4]. Hence, under the hypotheses of Theorem 1 we have

$$\iint_R \frac{\partial M}{\partial x}\, dxdy = \int_{J_0} M(x,y)\, dy = \int_{\Gamma_0} M(x,y)\, dy, \tag{7}$$

where it is to be emphasized that the right-hand integral of (7) is along the path curve Γ_0. It remains to be shown that the line integral along Γ_0 is equal to the line integral along Γ.[5]

Let $x = X(s)$, $y = Y(s)$, $(0 \leqq s \leqq L)$, denote the equations of Γ, where s is the arc length along this curve measured from B. To each $\epsilon > 0$ there corresponds a $\delta_\epsilon > 0$ such that if $B = p_0, p_1, \cdots, p_k = A$, $p_i: (x_i, y_i) = (X(s_i), Y(s_i))$, $(0 = s_0 < s_1 < \cdots < s_k = L)$, the length of each sub-arc $p_{i-1}p_i$ is less than δ_ϵ, and $(\xi_i, \eta_i) = (X(s'_i), Y(s'_i))$ is an arbitrary point with $s_{i-1} \leqq s'_i \leqq s_i$, then

$$\Big| \int_\Gamma M(x,y)\, dy - \sum_{i=1}^{k} M(\xi_i, \eta_i)(y_i - y_{i-1}) \Big| < \epsilon.$$

Similarly, for each $\epsilon > 0$ there exists a $\delta_{1\epsilon} > 0$ such that if $B = p'_0, p'_1, \cdots, p'_r = A$, $p'_j: (x'_j, y'_j) = (x_0(t_j), y_0(t_j))$, $(0 = t_0 < t_1 < \cdots < t_r = 1)$, the length of each path sub-arc $p'_{j-1}p'_j: x = x_0(t), y = y_0(t), (t_{j-1} \leqq t \leqq t_j)$, of Γ_0 is less than $\delta_{1\epsilon}$, and $(\xi'_j, \eta'_j) = (x_0(t'_j), y_0(t'_j))$ is an arbitrary point with $t_{j-1} \leqq t'_j \leqq t_j$, then

$$\Big| \int_{\Gamma_0} M(x,y)\, dy - \sum_{j=1}^{r} M(\xi'_j, \eta'_j)(y'_j - y'_{j-1}) \Big| < \epsilon.$$

Now let $B = \pi_0, \pi_1, \cdots, \pi_\kappa = A$ denote the distinct points of the combined sets $(p_0, \cdots, p_k)$, $(p'_0, \cdots, p'_r)$ arranged in order of occurrence along Γ; that is, $\pi_\alpha: (x_\alpha, y_\alpha) = (X(\sigma_\alpha), Y(\sigma_\alpha))$, $(\alpha = 0, 1, \cdots, \kappa)$, and $0 = \sigma_0 < \sigma_1 < \cdots < \sigma_\kappa = L$. Clearly the length of each sub-arc $\pi_{\alpha-1}\pi_\alpha$ of Γ

[5] Another method of proof of this result is as follows. If $s(t)$ denote the length of the simple sub-arc of Γ having end-points $B: (x_0(0), y_0(0))$ and $(x_0(t), y_0(t))$, then $s(t)$ may be shown to be absolutely continuous, in fact Lipschitzian, on $0 \leqq t \leqq 1$. The equality of the line integrals along Γ_0 and Γ is then a consequence of a general theorem on the change of variable in a simple integral given by Carathéodory [2], p. 563. From the standpoint of simplicity, however, the elementary proof given above seems preferable.

is less than δ_ϵ. With the sub-arc $\pi_{a-1}\pi_a$ we shall associate a definite point $(\xi_a, \eta_a) = (X(\sigma'_a), Y(\sigma'_a))$ with $\sigma_{a-1} \leqq \sigma'_a \leqq \sigma_a$. We shall now determine on Γ_0 a set of points $p_0'' = p_0' = B$, $p_1'', \cdots, p_\rho'' = p_r' = A$, p_β'': $(x_\beta'', y_\beta'') = (x_0(\tau_\beta), y_0(\tau_\beta))$, $(\beta = 0, 1, \cdots, \rho)$, $0 = \tau_0 < \tau_1 < \cdots < \tau_\rho = 1$, each of which is coincident with one of the points π_a, as follows. Suppose $p_0'', \cdots, p_\gamma''$ have been determined and that there is a value of j such that $\tau_\gamma = t_j$; that is, the point p_γ'' not only coincides geometrically with p_j', but they also correspond to the same parameter value of t. Then one or more succeeding points of the sequence $\{p_\beta''\}$ are defined by the following device. Let α, α' be the values such that geometrically $p'_j = \pi_a$, $p'_{j+1} = \pi_{a'}$. If $\alpha' - \alpha = 0$, 1 or -1, set $p''_{\gamma+1} = p'_{j+1} = \pi_{a'}$ and $\tau_{\gamma+1} = t_{j+1}$; if $\alpha' = \alpha + g$ or $\alpha - g$, $(g > 1)$, define $\tau_{\gamma+1}, \cdots, \tau_{\gamma+g}$ so that $t_j = \tau_\gamma < \tau_{\gamma+1} < \cdots < \tau_{\gamma+g} = t_{j+1}$ and correspondingly $(x_0(\tau_{\gamma+\nu}), y_0(\tau_{\gamma+\nu})) = (X(\sigma_{a+\nu}), Y(\sigma_{a+\nu}))$ or $(X(\sigma_{a-\nu}), Y(\sigma_{a-\nu}))$, $(\nu = 1, \cdots, g)$. This inductive definition clearly gives a unique sequence $B = p_0'', \cdots, p_\rho'' = A$ such that each p_β'' is one of the points π_a; moreover, consecutive points $p''_\beta, p''_{\beta+1}$ are either coincident or correspond to consecutive points $\pi_a, \pi_{a\pm1}$. If $p''_\beta = p''_{\beta+1}$, define $\tau'_\beta = \tau_\beta$, $(\xi''_\beta, \eta''_\beta) = (x_0(\tau'_\beta), y_0(\tau'_\beta))$. If $p''_\beta = \pi_a$, $p''_{\beta+1} = \pi_{a+1}$, let τ'_β denote a value on $\tau_\beta \leqq \tau'_\beta \leqq \tau_{\beta+1}$ such that $(\xi_\beta'', \eta_\beta'') \equiv (x_0(\tau_\beta'), y_0(\tau_\beta')) = (\xi_{a+1}, \eta_{a+1})$, this later point having been defined above. Correspondingly, if $p''_\beta = \pi_a$, $p''_{\beta+1} = \pi_{a-1}$, τ'_β is a parameter value on $\tau_\beta \leqq \tau_\beta' \leqq \tau_{\beta+1}$ such that $(\xi_\beta'', \eta_\beta'') \equiv (x_0(\tau_\beta'), y_0(\tau_\beta')) = (\xi_a, \eta_a)$. Clearly the partition $p_0'', \cdots, p_\rho''$ of Γ_0 is such that the length of each path sub-arc $p''_\beta p''_{\beta+1}$ is less than $\delta_{1\epsilon}$. It is also readily seen that the two sums

$$\sum_{a=1}^{k} M(\xi_a, \eta_a)(y_a - y_{a-1}), \qquad \sum_{\beta=1}^{\rho} M(\xi''_\beta, \eta''_\beta)(y''_\beta - y''_{\beta-1})$$

have the same value. Consequently,

$$\left| \int_\Gamma M(x, y)\,dy - \int_{\Gamma_0} M(x, y)\,dy \right| < 2\epsilon,$$

and since ϵ is an arbitrary positive number the line integral along Γ is equal to the integral along Γ_0.

By an argument similar to that above, or by the simple device of rotating the (x, y) coördinate axes through the angle $\pi/2$, one obtains the following result.

THEOREM 2. *Suppose that R is the interior of a simply closed rectifiable plane curve J, and that $N(x, y)$ is such that*: (i) *it is continuous on $R + J$*; (ii) *if $a \leqq x \leqq b$, $c \leqq y \leqq d$ is any rectangle in R, $N(x_0, y)$ is absolutely continuous in y on $c \leqq y \leqq d$ for x_0 almost everywhere on $a \leqq x_0 \leqq b$*; (iii) *$\partial N/\partial y$ is summable on R. Then*

$$\iint_R \frac{\partial N}{\partial y}\,dxdy = -\int_J N(x,y)\,dx, \tag{8}$$

the line integral being taken around J in the positive direction.

Combining the results of Theorems 1 and 2, we have the following general form of Green's lemma: *if $M(x,y)$ and $N(x,y)$ satisfy the conditions of the above theorems,*

$$\iint_R \left[\frac{\partial M}{\partial x} - \frac{\partial N}{\partial y}\right] dxdy = \int_J Mdy + Ndx. \tag{9}$$

THEOREM 3. *If $f(z)$ is holomorphic on the interior R of a simply closed rectifiable curve J, and continuous on $R + J$, then*

$$\int_J f(z)\,dz = 0.$$

It is readily proved (see, for example, [10]) that the integral of f around the boundary of any rectangle lying in R is equal to zero. From this one can readily deduce that the integral of f around any simply closed polygon lying in R, and each of whose component straight line segments is parallel to either the x- or y-axis, is also equal to zero. Hence the integral of f around each of the simply closed polygons J_n used in the proof of Theorem 1 is equal to zero, and passing to the limit it follows that the integral of f around the path curve J_0 is also zero. It then follows by the device used in Theorem 1 that the integral of f around J is zero. Clearly one may separate the line integral of f into its real and pure imaginary parts and follow the steps outlined above, or one may treat it throughout as a line integral involving the complex variable z.

Finally, we shall prove the following theorem which, under somewhat stronger conditions on J, has been announced by Hestenes [7]. This result is useful in the derivation of edge conditions for an extremizing surface in multiple integral problems of the calculus of variations.

THEOREM 4. *Suppose that $u(x,y)$, $v(x,y)$ and $w(x,y)$ are continuous on $R + J$, where R is the interior of the simply closed rectifiable curve J, and for an arbitrary $\eta(x,y)$ which is Lipschitzian on $R + J$, and identically zero on J, we have*[6]

---

[6] From the proof of this theorem it is seen that the class of functions η for which (10) is supposed to hold need not consist of all functions which are Lipschitzian on $R + J$, and identically zero on J. Indeed, the proof utilizes (10) only for functions η which are identically zero exterior to some rectangle $a \leqq x \leqq b$, $c \leqq y \leqq d$ lying in R.

$$\iint_R [u\eta_x + v\eta_y + w\eta]dxdy = 0. \tag{10}$$

Then

$$\iint_R [u\eta_x + v\eta_y + w\eta]dxdy = \int_J \eta[udy - vdx] \tag{11}$$

for every $\eta(x,y)$ *such that*: (i) η *is continuous on* $R+J$; (ii) *if* $a \leqq x \leqq b$, $c \leqq y \leqq d$ *is any rectangle in* R, $\eta(x_0, y)$ *is absolutely continuous in* y *on* $c \leqq y \leqq d$ *for* x_0 *almost everywhere on* $a \leqq x_0 \leqq b$, *and* $\eta(x, y_0)$ *is absolutely continuous in* x *on* $a \leqq x \leqq b$ *for* y_0 *almost everywhere on* $c \leqq y_0 \leqq d$; (iii) η_x *and* η_y *are summable on* R.

Let J' and R' denote, respectively, the boundary and interior of a rectangle $a \leqq x \leqq b, c \leqq y \leqq d$ lying in R, and suppose that η is chosen to be Lipschitzian on $R+J$, and identically zero outside J'. Then the conditions of the theorem imply

$$\iint_{R'} [u\eta_x + v\eta_y + w\eta]dxdy = 0. \tag{12}$$

Let h be so small that for $|s| \leqq h$, $|t| \leqq h$ the rectangle $a - s \leqq x \leqq b - s$, $c - t \leqq y \leqq d - t$ remains interior to J. Then $\eta(x-s, y-t)$ is Lipschitzian on $R+J$, and identically zero outside the associated rectangle; consequently the corresponding integral for this function is also zero. By a simple change of variable it follows that

$$\begin{aligned} 0 = \iint_{R'} [&u(x+s, y+t)\eta_x(x,y) + v(x+s, y+t)\eta_y(x,y) \\ &+ w(x+s, y+t)\eta(x,y)]dxdy \end{aligned} \tag{13}$$

for arbitrary s, t satisfying $|s| \leqq h$, $|t| \leqq h$. Set

$$u^{(h)}(x,y) = \frac{1}{4h^2}\int_{-h}^{h}\int_{-h}^{h} u(x+s, y+t)dsdt;$$

$v^{(h)}(x,y)$ and $w^{(h)}(x,y)$ are defined by corresponding integral means. The functions $u^{(h)}$, $v^{(h)}$, $w^{(h)}$ are seen to be of class C^1 on $R'+J'$; moreover, as $h \to 0$ they tend uniformly on $R'+J'$ to u, v, w, respectively. By integrating (13) with respect to s, t, it follows that

---

The proof is still valid if we merely assume that (10) holds for functions of this latter sort which, in addition, may be further restricted to have continuous derivatives in R of a prescribed arbitrarily high order.

$$(14)\qquad 0=\iint_{R'} [u^{(h)}\eta_x+v^{(h)}\eta_y+w^{(h)}\eta]dxdy$$
$$=\iint_{R'} [w^{(h)}-\frac{\partial}{\partial x}u^{(h)}-\frac{\partial}{\partial y}v^{(h)}]\eta\, dxdy$$

for arbitrary functions η which are Lipschitzian on $R'+J'$ and vanish on J'. Because of the arbitrariness of η, we deduce that $w^{(h)}-\partial u^{(h)}/\partial x-\partial v^{(h)}/\partial y \equiv 0$ on $R'+J'$. The above demonstration is really a proof of the integral form of the Haar equations associated with the double integral (10), where the path of integration is the boundary of a rectangle interior to J.

For an arbitrary η satisfying conditions (i), (ii), (iii) of Theorem 4, it follows from Green's lemma for a rectangle that

$$(15)\qquad \iint_{R'} [u^{(h)}\eta_x+v^{(h)}\eta_y+w^{(h)}\eta]dxdy=\int_{J'}\eta[u^{(h)}dy-v^{(h)}dx].$$

Letting $h\to 0$ in (15), we obtain the same relation with u, v and w replacing their corresponding integral means; that is, we have relation (11) with R, J replaced by R', J', respectively. Since R' is the interior of an arbitrary rectangle lying in R, one may then deduce that the corresponding relation is valid for each of the simply closed polygons J_n with interior R_n appearing in the proof of Theorem 1. The conditions (i), (ii) and (iii) of Theorem 4 then enable us to pass from the polygons J_n to the simply closed curve J as in the proof of Theorem 1, and thus obtain the desired result (11).

The results of this section may be readily extended to the case of a bounded region R whose complete boundary J consists of a finite number of disjoint simply closed rectifiable curves.

3. Proof of Lemma I. Let K be a circle enclosing the simple closed curve J and its interior, and denote by Π a simple polygonal arc joining O to a point T of K which lies, except for its end-points, exterior to J and interior to K. Such an arc clearly exists since by the Jordan curve theorem the exterior of J is connected, while O is accessible from arbitrary points which are exterior to J, and which belong to a sufficiently small neighborhood of O, by straight line segments which belong, except for O, to the exterior of J. Let d_0 denote the distance between the point set $\Pi+K$ and the arc $\Gamma: BCA$. For a given $\epsilon>0$, let $\epsilon_1=\mathrm{Min}\{\epsilon, d_0\}$. Then there exists an integer N such that $(5L\sqrt{2})/(4N)<\epsilon_1$, where L is the length of the arc Γ of J. Divide Γ into N successive arcs $\gamma_1,\cdots,\gamma_N$ each of length L/N, and denote by p_i the mid-point, in the sense of length along Γ, of the arc γ_i, $(i=1,\cdots,N)$. Now let Δ_i denote the perimeter of the square with center p_i, length of side equal to

$(5L)/(4N)$, and each of whose sides is parallel to either the x- or y-axis. The arc γ_i is seen to be interior to Δ_i. Clearly no one of the perimeters Δ_i intersects or encloses a point of the set $\Pi + K$; moreover, each point of one of the Δ_i is in an ϵ-neighborhood of the arc Γ. The length of any simple polygonal arc made up of pieces of the sides of the perimeters Δ_i is seen not to exceed $5L$. Hence the result of Lemma I is established if it can be shown that there exists a particular such arc $\Gamma': B'C'A'$ satisfying conditions (a) and (b) of the lemma.

There are various general theorems of point set theory from which this desired conclusion is readily deducible. Because of the particular character of the curves Δ_i, however, one may give a proof of this result which involves only such elementary properties of simply closed curves and simple arcs as those stated in Theorems 3–7 of Moore [12] (Theorems 24–27 and 29 of Moore [11]).

We begin by proving the following result.

Auxiliary Lemma. *There exists a simply closed polygon P such that: (α) every point of P belongs to some one of the perimeters Δ_i; (β) the interior of P contains the interiors of all of the Δ_i, $(i = 1, \cdots, N)$.*

Since the sum of the interiors of the perimeters Δ_i is a connected set, this result in turn is a very special case of a theorem given by R. L. Moore ([11], p. 156, or [12], p. 166) for general simply closed curves. Because of the special character of the curves Δ_i, however, a simplified proof of this result can be given. We shall prove by induction that for each k, $(k = 1, \cdots, N)$, there exists a simply closed polygon P_k satisfying conditions (α), (β) with respect to $\Delta_1, \cdots, \Delta_k$; then $P = P_N$ is a polygon of the desired type. Clearly for $k = 1$, $P_1 = \Delta_1$ itself. Now suppose that there exists a simply closed polygon P_k satisfying (α), (β) with respect to $\Delta_1, \cdots, \Delta_k$. If I_k and I denote the interior of P_k and Δ_{k+1}, respectively, then I_k and I have points in common in view of the definition of the perimeters Δ_i. If I is a subset of I_k, then clearly we might choose $P_{k+1} = P_k$ for $\Delta_1, \cdots, \Delta_{k+1}$; whereas, if I_k is a subset of I, set $P_{k+1} = \Delta_{k+1}$. Consequently, suppose that neither I nor I_k is a subset of the other. Then there exist arcs of Δ_{k+1} which lie, except for their end-points, exterior to P_k, while the end-points belong to P_k. Since P_k consists of a finite number of straight line segments each of which is parallel to one of the coördinate axes, there can exist only a finite number of such arcs of Δ_{k+1}. Let $\delta_1: a_1c_1b_1, \cdots, \delta_r: a_rc_rb_r$ denote the totality of such arcs. If $a_1d_1b_1$ and $a_1d'_1b_1$ denote the two simple arcs of P_k with end-points a_1 and b_1 then one of these arcs, say $a_1d_1b_1$ for definiteness, is such that the interior of P_k, as well as all the interior points of the other arc $a_1d'_1b_1$, belong to the interior

of the simply closed polygon $P_{k,1}: a_1c_1b_1d_1a_1$. Let j_1 be the first value of j such that the interior points of the arc δ_{j_1} belong to the exterior of $P_{k,1}$. If $a_{j_1}d_2b_{j_1}$ and $a_{j_1}d'_2b_{j_1}$ denote the two simple arcs of $P_{k,1}$ with end-points a_{j_1} and b_{j_1} then one of these arcs, say $a_{j_1}d_2b_{j_1}$, is such that the interior of $P_{k,1}$, as well as all the interior points of the other arc $a_{j_1}d'_2b_{j_1}$, belong to the interior of the simply closed polygon $P_{k,2}: a_{j_1}c_{j_1}b_{j_1}d_2a_{j_1}$. After g steps, $1 \leqq g \leqq r$, one obtains a simply closed polygon $P_{k,g}$ such that every point of the arcs $\delta_1, \cdots, \delta_r$ either belongs to $P_{k,g}$ or is interior to this polygon. Since the exterior of $P_{k,g}$ is connected it follows that all points interior to Δ_{k+1}, as well as those interior to P_k, belong to the interior of $P_{k,g}$; that is, $P_{k+1} = P_{k,g}$ satisfies conditions (α), (β) for $\Delta_1, \cdots, \Delta_{k+1}$. Hence by induction one arrives at the result of the above auxiliary lemma.

Returning to the proof of Lemma I, it is first noted that the polygon P determined by the auxiliary lemma must have points belonging to the interior of J. For otherwise, since the interior of J is connected, it would follow that O could be joined to a point C' of Γ by a simple arc which lies, except for O and C', interior to J and consequently not intersecting P. This, however, is a contradiction since O is exterior to P while the arc Γ is interior to this polygon. If p is a point of P interior to J, then this point is seen to belong to an arc acb of P which lies, except for its end-points, interior to J. Moreover, in view of the character of P, each of the end-points a and b is either interior to AO or interior to OB. Since each line component of P is parallel to either the x- or y-axis there are only a finite number of such arcs acb; let $\delta_j: a_jc_jb_j$, $(j = 1, \cdots, l_0)$, denote the totality of such simple polygonal arcs. If the conclusion of the lemma is not true, then for each such arc both the end-points a_j, b_j belong to either OB or AO. Now for a particular j_0 it may be that the arc δ_{j_0}, except for its end-points, is interior to the simply closed curve formed by the arcs a_jb_j and $b_jc_ja_j$ corresponding to a value $j \neq j_0$. All such arcs δ_{j_0} will be omitted; suppose that there remain l other arcs $a_jc_jb_j$. This collection of arcs may then be ordered, with a possible interchange of symbols a_j and b_j, so that on the segment AB the points $A, a_1, b_1, \cdots, a_r, b_r, O, a_{r+1}, b_{r+1}, \cdots, a_l, b_l, B$ appear in the order written. Using the elementary properties of simply closed curves and simple arcs referred to above, it may be readily shown that

$$Aa_1c_1b_1 \cdots a_rc_rb_rOa_{r+1}c_{r+1}b_{r+1} \cdots a_lc_lb_lBCA$$

is a simply closed curve $\mathcal{J}$ whose interior belongs to the interior of J, and is such that no point of P is interior to $\mathcal{J}$. It would again be true that O could be joined to a point C' of Γ by a simple arc which lies, except for O and C',

interior to $\mathcal{J}$, which again is a contradiction to the fact that P separates O from the arc Γ. Hence, one of the arcs $a_j c_j b_j$ must have one end-point interior to AO, and the other end-point interior to OB; such an arc satisfies the conditions of the lemma.

THE UNIVERSITY OF CHICAGO,
CHICAGO, ILLINOIS.

BIBLIOGRAPHY

1. H. E. Bray, "Green's lemma," *Annals of Mathematics*, vol. 26 (1924-5), pp. 278-286.

2. C. Carathéodory, *Vorlesungen über reelle Funktionen*, Teubner, second edition, 1927.

3. E. Goursat and E. R. Hedrick, *Mathematical Analysis*, vol. II—part I. *Functions of a complex variable*, Ginn and Co., 1916.

4. L. M. Graves, "On the existence of the absolute minimum in space problems of the calculus of variations," *Annals of Mathematics*, vol. 28 (1927), pp. 153-170.

5. W. Gross, "Das isoperimetrische Problem bei Doppelintegralen," *Monatshefte für Mathematik und Physik*, vol. 27 (1916), pp. 70-120.

6. H. Heilbronn, "Zu dem Integralsatz von Cauchy," *Mathematische Zeitschrift*, vol. 37 (1933), pp. 37-38.

7. M. R. Hestenes, "On the first necessary condition for minima of double integrals," for abstract see *Bulletin of the American Mathematical Society*, vol. 45 (1939), p. 519.

8. E. W. Hobson, *The theory of functions of a real variable*, vol. 2, second edition, Cambridge, 1926.

9. E. Kamke, "Zu dem Integralsatz von Cauchy," *Mathematische Zeitschrift*, vol. 35 (1932), pp. 539-543.

10. E. H. Moore, "A simple proof of the fundamental Cauchy-Goursat theorem," *Transactions of the American Mathematical Society*, vol. 1 (1900), pp. 499-506.

11. R. L. Moore, "On the foundations of plane analysis situs," *Transactions of the American Mathematical Society*, vol. 17 (1916), pp. 131-164.

12. R. L. Moore, "Foundations of point set theory," *American Mathematical Society Colloquium Publications*, vol. 13, New York, 1932.

13. S. Pollard, "On the conditions for Cauchy's Theorem," *Proceedings of the London Mathematical Society*, vol. 21 (1923), pp. 456-482.

14. E. B. Van Vleck, "An extension of Green's lemma to the case of a rectifiable boundary," *Annals of Mathematics*, vol. 22 (1920-21), pp. 226-237.

ON THE ASYMPTOTIC BEHAVIOR OF THE RIEMANN ZETA-FUNCTION ON THE LINE $\sigma = 1$.*

By Aurel Wintner.

Littlewood has shown [1] that, on Riemann's hypothesis,

$$(1) \qquad 1/\zeta(1+it) = O(\log\log t)\,;\ \text{while}\ 1/\zeta(1+it) \neq o(\log\log t)$$

is true, according to Bohr and Landau, without any hypothesis.[2] Without any hypothesis, Littlewood found [3]

$$(2) \qquad 1/\zeta(1+it) = O(\log t/\log\log t).$$

The proof of (2) depends on very much more than on Hadamard's theorem

$$(3) \qquad \zeta(1+it) \neq 0.$$

Since (3) is, according to Ikehara's Tauberian theorem, equivalent to the prime number theorem, it is natural to ask *what relative amount of time does* $\zeta(1+it)$ *spend in the immediate neighborhood of the critical value* $\zeta = 0$. In contrast with the problem of estimates, as represented by (1) and (2), this question disregards arbitrarily distant t-ranges on which $\zeta(1+it)$ is very small but which are of relative measure 0.

In this direction of asymptotic distributions, the method of infinite convolutions supplied,[4] among other things, the following information: $r = |\zeta(1+it)|$ has an asymptotic distribution function $\psi(r)$, $0 \leqq r < \infty$, which has derivatives of arbitrarily high order for every $r \geqq 0$, and a density which is positive for every $r > 0$ but vanishes, together with all higher derivatives, at $r = 0$, since

$$(4) \qquad \psi(r) = O\exp(-\lambda\log^2 r), \qquad r \to 0,$$

* Received December 2, 1940.

[1] J. E. Littlewood, "On the Riemann zeta-function," *Proceedings of the London Mathematical Society*, ser. 2, vol. 24 (1924), pp. 175-201; cf. "Mathematical Notes (5): On the function $1/\zeta(1+ti)$," *ibid.*, vol. 27 (1927), pp. 349-357.

[2] H. Bohr and E. Landau, "Nachtrag zu unseren Abhandlungen aus den Jahrgängen 1910 und 1923," *Göttinger Nachrichten*, 1924, pp. 168-172.

[3] J. E. Littlewood, "Researches in the theory of the Riemann ζ-function," *Proceedings of the London Mathematical Society*, ser. 2, vol. 20 (1922), Records, pp. xxii-xxviii.

[4] B. Jessen and A. Wintner, "Distribution functions and the Riemann zeta function," *Transactions of the American Mathematical Society*, vol. 38 (1935), pp. 48-88, more particularly, § 10 and § 14.

holds for every fixed $\lambda > 0$. It is understood that the function ψ of r is defined by the property that

$$\lim_{T\to\infty} \frac{\tau_r(T)}{2T} = \psi(r) \tag{5}$$

holds for every r, where $\tau_r(T)$ denotes the sum of the lengths of those subintervals of the interval $-T \leqq t \leqq T$ on which $|\zeta(1+it)| \leqq r$.

The estimate (4) depends on those properties of $|\zeta(1+it)|$ (or, rather, of $\log |\zeta(1+it)|$) which make possible the application of the theory of infinite convolutions.[5] On the other hand, the proof of (4) did not use (3), and so there arises the question as to a possible refinement of (4) by means of the prime number theorem or its known remainder terms. This question is the more natural as nothing better and nothing worse than (4) was established [6] on *any* fixed line σ *within* the strip $\frac{1}{2} < \sigma < 1$.

The purpose of the following considerations is to answer this question. It turns out that, by using deeper properties of $\zeta(s)$, the estimate (4) of the asymptotic distribution function of $|\zeta(1+it)|$ can be greatly improved, as follows:

$$\psi(r) = O \exp(-\lambda/r), \qquad r \to 0; \tag{6}$$

λ being an arbitrarily fixed positive number.

The proof of this Gaussian estimate will depend on a property of $\zeta(s)$ somewhat more recondite than the property (3) (which is sufficient for the prime number theorem). The property of $\zeta(s)$ in question is [7] that the *formal trigonometric series*

$$1/\zeta(1+it) = \sum_{n=1}^{\infty} \mu(n) n^{-1} \exp(-it \log n), \tag{7}$$

a series which is convergent for $-\infty < t < \infty$ in virtue of (3) and of M. Riesz' Tauberian theorem, is the *Fourier series* (B^2) of the function $1/\zeta(1+it)$; so that $1/\zeta(1+it)$ is almost periodic (B^2) and

$$1/\zeta(1+it) \sim \sum_{n=1}^{\infty} \mu(n) n^{-1} \exp(-it \log n). \tag{8}$$

This fact will make possible the adaptation of a method recently applied to the explicit formula of the prime number theory.[8]

---

[5] *Loc. cit.*[4]

[6] *Ibid.*

[7] A. Wintner, "The almost periodic behavior of the function $1/\zeta(1+it)$," *Duke Mathematical Journal*, vol. 2 (1936), pp. 443-446.

[8] A. Wintner, "On the distribution function of the remainder term of the prime number theorem," *American Journal of Mathematics*, vol. 63 (1941), pp. 233-248.

The proof of (8) given *loc. cit.* uses somewhat more than the prime number theorem, since the non-vanishing of $\zeta(s)$ is needed not only on the line $\sigma = 1$ but also in a certain domain on the left of $\sigma = 1$. However, the domain in question is in the interior even of the de la Vallée Poussin domain; so that (8) is independent of any hypothesis. Actually, it is not known what, if anything, follows in Tauberian terms for the distribution of primes, if (7) is replaced by (8); so that it is an open question whether or not (8) can be proved without a use of (3), at least.

At any rate, (6) will be proved without Riemann's hypothesis. Incidentally, the method to be applied would not improve on the Gaussian estimate (6) even under Riemann's hypothesis. Furthermore, it does not seem to be easy to transfer (8) from $\sigma = 1$ to $\frac{1}{2} < \sigma < 1$ by using Riemann's hypothesis.

It was shown *loc. cit.*[4] that, corresponding to (4),

$$\text{(4 bis)} \qquad 1 - \psi(r) = O \exp(-\lambda \log^2 r); \qquad r \to \infty.$$

It will be clear from the proof of (6) that (4 bis) can be improved to

$$\text{(6 bis)} \qquad 1 - \psi(r) = O \exp(-\lambda r), \qquad r \to \infty,$$

since (8) is paralleled by

$$\text{(8 bis)} \qquad \zeta(1 + it) \sim \sum_{n=1}^{\infty} n^{-1} \exp(-it \log n)$$

(actually, it is easy to prove that (8 bis) holds, without any hypothesis,[9] not only in the sense (B^2) but also in the sense (B^p), where p is arbitrarily large). Needless to say, (8 bis) is, in contrast with (8), in no way connected with (3). Correspondingly, (6) is much deeper than (6 bis). This agrees with the fact that, on the one hand, (6) and (6 bis) express estimates of the asymptotic densities of large values of $1/\zeta(1 + it)$ and $\zeta(1 + it)$, respectively, and that, on the other hand, the classical (O, Ω)-problems are known[10] to be much easier for $\zeta(1 + it)$ than for $1/\zeta(1 + it)$.

Since the space of the functions which are almost periodic (B^q) for a fixed $q \geqq 1$ is a complete space, the proof of the Young-Hausdorff extension of the Fischer-Riesz existence theorem can be transcribed from (L^q) to (B^q) without any change.[11] Accordingly, if $\{\lambda_n\}$ is any sequence of distinct real numbers, and if $\{a_n\}$ is any sequence of numbers such that

---

[9] In this connection, cf. A. Wintner, "Riemann's hypothesis and almost periodic behavior," *Universidad Mayor de San Marcos*, Lima (Peru), no. 430, vol. 41 (1939), pp. 575-585.

[10] A comprehensive account of these problems may be found in Titchmarsh's Tract.

[11] H. R. Pitt, "On the Fourier coefficients of almost periodic functions," *Journal of the London Mathematical Society*, vol. 14 (1939), pp. 143-150.

$$\sum_{n=1}^{\infty} | a_n |^{q/(q-1)} < \infty$$

holds for a fixed $q > 1$, then there exists a function $f(t)$, $-\infty < t < \infty$, which is almost periodic (B^q) and such that

$$f(t) \sim \sum_{n=1}^{\infty} a_n \exp(i\lambda_n t);$$

furthermore,

$$[M\{| f |^q\}]^{1/q} \leqq [\sum_{n=1}^{\infty} | a_n |^{q/(q-1)}]^{(q-1)/q},$$

where

$$M\{g\} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} g(t)dt.$$

If $a_n = \mu(n)n^{-1}$, then, since $| \mu(n) | \leqq 1$, the condition imposed on $\{a_n\}$ is satisfied by every $q > 1$. Hence, on choosing

$$a_n = \mu(n)n^{-1}, \qquad \lambda_n = -\log n, \qquad q = 2k,$$

where k is any fixed positive integer, one sees that there exists for every k a function $f = f_{2k}$ which is almost periodic (B^{2k}) and such that

$$(9) \qquad f_{2k}(t) \sim \sum_{n=1}^{\infty} \mu(n)n^{-1} \exp(-it \log n);$$

furthermore,

$$[M\{| f_{2k} |^{2k}\}]^{1/2k} \leqq [\sum_{n=1}^{\infty} | \mu(n)n^{-1} |^{2k/(2k-1)}]^{(2k-1)/2k},$$

and so, since $\sum_{n=1}^{\infty} | \mu(n)n^{-1} |^{2k/(2k-1)} < \sum_{n=1}^{\infty} n^{-2k/(2k-1)} = \zeta\left(\frac{2k}{2k-1}\right) > 1$,

$$(10) \qquad [M\{| f_{2k} |^{2k}\}]^{1/2k} < \zeta\left(\frac{2k}{2k-1}\right) < \infty; \qquad k = 1, 2, \cdots.$$

Since the Fourier series (9) is one and the same for every k, and since the n-th partial of the series (9) tends in the mean of (B^{2k}) to $f_{2k}(t)$, the asymptotic distribution of $f_{2k}(t)$ is independent of the subscript.[12] Hence, the non-negative function $| f_{2k}(t) |$, which is almost periodic (B^{2k}), has an asymptotic distribution function

$$(11) \qquad \phi = \phi(r), \qquad 0 \leqq r < \infty, \qquad (\phi(r) \equiv 0 \text{ for } -\infty < r \leqq 0)$$

which is independent of k. Consequently,[12]

---

[12] Cf. B. Jessen and A. Wintner, *loc. cit.*[4], p. 76.

(12) $$M_{2k}(\phi) = M\{|f_{2k}|^{2k}\} < \infty; \qquad k = 1, 2, \cdots,$$

where $M_k(\phi)$ denotes the k-th momentum,

(13) $$0 < M_k(\phi) = \int_0^\infty r^k d\phi(r),$$

of the common asymptotic distribution function of the absolute values of the functions $f_2(t), f_4(t), f_6(t), \cdots$, which are all almost periodic (B^2).

On the other hand, it is seen from (8) and (9) that the function $1/\zeta(1+it)$, which is almost periodic (B^2), has the same Fourier series, and therefore the same asymptotic distribution function, as the function $f_{2k}(t)$. Thus the common asymptotic distribution function, (11), of the functions $|f_{2k}(t)|$ is the asymptotic distribution function of $1/|\zeta(1+it)|$ also. Hence, $\phi(r)$ has, for every r, derivatives of arbitrarily high order and the density, that is, first derivative, is such that

(14) $$\phi'(r) > 0 \text{ for } 0 < r < \infty, \text{ while } \phi'(0) = 0; \qquad (\phi' = d\phi/dr).$$

In particular, (13) reduces to

(15) $$M_k(\phi) = \int_0^\infty r^k\phi'(r)dr, \text{ where } k = 0, 1, 2, \cdots; \qquad M_0(\phi) = 1$$

The order of magnitude of the momenta $M_k(\phi)$ for large k can be estimated by

(16) $$[M_{2k}(\phi)]^{1/2k} < Ck; \qquad k = 1, 2, \cdots,$$

where $C > 0$ is a numerical constant. In fact, since

$$\zeta(1+\epsilon) \sim 1/\epsilon \text{ as } \epsilon \to +0,$$

there exists a $C > 0$ for which

$$\zeta\left(\frac{2k}{2k-1}\right) < Ck; \qquad k = 1, 2, \cdots.$$

Hence, (16) follows from (10) and (12).

Actually, (16) can be replaced by

(17) $$[M_{2k}(\phi)]^{1/2k} = o(k); \qquad k \to \infty.$$

In order to see this, it is sufficient to choose a fixed positive integer m; to replace, for this fixed m, the functions (9) by the functions $f_{2k}^{(m)}$ of class (B^{2k}) which are defined by

$$(9_m) \qquad f_{2k}^{(m)}(t) \sim \sum_{n=m}^{\infty} \mu(n) n^{-1} \exp(-it \log n)$$

(so that $f_{2k}^{(1)} = f_{2k}$); to repeat, for the sequence $f_2^{(m)}, f_4^{(m)} f_6^{(m)}, \cdots$ and for the fixed value of m, everything that led to (17) in case of the sequence $f_2 = f_2^{(1)}, f_4 = f_4^{(1)}, f_6 = f_6^{(1)}, \cdots$; finally, to let $m \to \infty$.

The balance of the proof of the desired Gaussian estimate of the asymptotic distribution can now be carried out by using standard devices, as follows:

According to Stirling,

$$k!^{1/k} \sim e^{-1}k; \qquad k \to \infty.$$

Consequently, (17) can be written in the form

$$[M_{2k}(\phi)/2k!]^{1/2k} \to 0; \qquad k \to \infty.$$

Thus, if λ denotes a positive parameter,

$$\sum_{k=0}^{\infty} \lambda^k M_k(\phi)/k! < \infty \text{ for every } \lambda.$$

Accordingly, from (13),

$$\int_0^{\infty} \exp(\lambda r) d\phi(r) < \infty \text{ for every } \lambda.$$

Since λ is arbitrary, it follows that

$$\phi(r) = O \exp(-\lambda r), \qquad r \to \infty,$$

holds for every fixed $\lambda > 0$. Hence, in order to complete the proof of (6), it is sufficient to observe that $\psi(r)$ and $\phi(r)$ denote the asymptotic distribution functions of $|\zeta(1+it)|$ and $1/|\zeta(1+it)|$, respectively; distribution functions possessing densities which are continuous and positive for $0 < r < \infty$ and tend to 0 when either $r \to 0$ or $r \to \infty$; cf. the beginning of this paper.

It is of methodical interest[13] that (6), (6 bis) do, while (4), (4 bis) did not, insure that the corresponding Stieltjes momentum problems are of the determined type.

THE JOHNS HOPKINS UNIVERSITY.

[13] Cf. B. Jessen and A. Wintner, *loc. cit.*[4], p. 76.

ON THE CONVEXITY OF AVERAGES OF ANALYTIC ALMOST PERIODIC FUNCTIONS.*

By Philip Hartman and Aurel Wintner

Let $f(\sigma, t)$ be either a regular analytic or a regular harmonic function in the strip $\gamma < \sigma < \delta$; $-\infty < t < +\infty$, and suppose that $f(\sigma, t)$ is uniformly almost periodic in this strip (for instance, let the function be a uniformly convergent Dirichlet series or its real part). Let p be any fixed index not less than 1, and put

$$F(\sigma) = M_t\{|f(\sigma + it)|^p\},$$

where

$$M_t\{h(t)\} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} h(t)\,dt.$$

In a problem connected with Riemann's zeta-function, use had to be made of a lemma which states that $F(\sigma)$ satisfies the maximum principle[1]; that is, that if $[\alpha, \beta]$ is any closed interval contained in the open interval (γ, δ), the maximum of $F(\sigma)$ on the interval $[\alpha, \beta]$ is attained either at $\sigma = \alpha$ or at $\sigma = \beta$.

This lemma will now be refined by proving that *$F(\sigma)$ is a convex function on the interval (γ, δ).*

The result can be thought of as the almost periodic extension of the three circle theorems of the Hadamard-Hardy type and their generalizations. In fact, the classical three circle theorems are obtained by particularizing the condition of almost periodicity to the periodicity condition $f(\sigma, t) \equiv f(\sigma, t + 2\pi)$; in which case the t-average over the line σ reduces to the ordinary mean value over the circle of radius $r = e^{-\sigma}$.

Actually, the almost periodic function $f(\sigma, t)$ will be subjected only to a subharmonic assumption; furthermore, it will not be necessary to assume that the almost periodicity is uniform.

If a function $g = g(\sigma, t)$, defined on a strip

$$\alpha \leqq \sigma \leqq \beta, \qquad -\infty < t < +\infty,$$

* Received December 30, 1940.

[1] A. Wintner, "On the distribution function of the remainder term of the prime number theorem," *American Journal of Mathematics*, vol. 63 (1941), pp. 233-248.

is non-negative, subharmonic and such as to satisfy, uniformly in σ, an estimate of the form

$$g(\sigma, t) = O(|t|^C) \text{ as } t \to \pm\infty,$$

where C is a sufficiently large constant, and if

$$\mu(T; \tau; \sigma) = \frac{1}{2T}\int_{\tau-T}^{\tau+T} g(\sigma, t)\,dt$$

tends, for every fixed σ in the interval $\alpha \leqq \sigma \leqq \beta$, to a limit $M_t\{g(\sigma, t)\}$ uniformly in τ as $T \to \infty$, then the function $M_t\{g(\sigma, t)\}$ of σ is convex on the interval $\alpha \leqq \sigma \leqq \beta$.

If $M_t\{g(\sigma, t)\}$ is a convex function of σ on every sufficiently short subinterval of $\alpha \leqq \sigma \leqq \beta$, it is a convex function on the whole interval $\alpha \leqq \sigma \leqq \beta$. Hence, it can be assumed (after a translation) that $-\frac{1}{2}\pi < \alpha$ and $\beta < \frac{1}{2}\pi$. Then, if σ is any point of the interval $\alpha \leqq \sigma \leqq \beta$, the subharmonic character of g and the uniform $O(|t|^C)$-estimate (and, as a matter of fact, even a weaker estimate of the Phragmén-Lindelöf type) are known [2] to imply that there exists, for every real number γ, for every $\epsilon > 0$, and for every $T > 0$, a real number

$$t_0 = t_0(\gamma, \epsilon, \sigma, T)$$

such that

$$\int_{-T}^{T} g(\sigma, t)\,dt - \gamma\sigma - \epsilon(e^T - e^{-T})\cos\sigma$$
$$\leqq \operatorname{Max}\left(\int_{-T}^{T} g(\alpha, t + t_0)\,dt - \gamma\alpha,\ \int_{-T}^{T} g(\beta, t + t_0)\,dt - \gamma\beta\right).$$

Let c be any real number and choose the real number γ, mentioned before, to be equal to $2Tc$. Then, on dividing the preceding inequality by $2T$, one sees that, if σ is arbitrarily fixed in the strip $\alpha \leqq \sigma \leqq \beta$, there exists, for every real number c, for every $\epsilon > 0$, and for every $T > 0$, a real number

$$t^0 = t^0(c, \epsilon, \sigma, T)$$

such that

$$\mu(T; 0; \sigma) - c\sigma - \tfrac{1}{2}\epsilon T^{-1}(e^T - e^{-T})\cos\sigma$$
$$\leqq \operatorname{Max}\left(\mu(T; t^0; \alpha) - c\alpha,\ \mu(T; t^0; \beta) - c\beta\right).$$

[2] Cf. G. H. Hardy, A. E. Ingham and G. Pólya, "Notes on moduli and mean values," *Proceedings of the London Mathematical Society*, Ser. 2, vol. 27 (1928), pp. 401-409 (more particularly pp. 407-408), where further references are given

For a given $\eta > 0$, choose a positive $R = R(\eta)$ so large that

$$|\mu(T;t^0;\alpha) - M_t\{g(\alpha,t)\}| < \eta \text{ and } |\mu(T;t^0;\beta) - M_t\{g(\beta,t)\}| < \eta$$

whenever $T > R(\eta)$. The existence of such a function R of η alone is assured by the assumption that the limit relation

$$\mu(T;\tau;\sigma) \to M_t\{g(\sigma,t)\}, \qquad T \to \infty,$$

holds uniformly for all values of τ (where $\tau = t^0(c,\epsilon,\sigma T) = t^0$), where $\sigma(=\alpha,\beta)$ is fixed. Thus,

$$\mu(T;0;\sigma) - c\sigma - \tfrac{1}{2}\epsilon T^{-1}(e^T - e^{-T})\cos\sigma$$
$$\leqq \eta + \operatorname{Max}\,(M_t\{g(\alpha,t)\} - c\alpha,\ M_t\{g(\beta,t)\} - c\beta),$$

whenever $T > R(\eta)$.

Thus far c has been arbitrary. Determine now c by the condition

$$M_t\{g(\alpha,t)\} - c\alpha = M_t\{g(\beta,t)\} - c\beta.$$

Then the preceding inequality can be written in the form

$$\mu(T;0;\sigma) - \tfrac{1}{2}\epsilon T^{-1}(e^T - e^{-T})\cos\sigma$$
$$\leqq \eta + \frac{\beta-\sigma}{\beta-\alpha}M_t\{g(\alpha,t)\} + \frac{\sigma-\alpha}{\beta-\alpha}M_t\{g(\beta,t)\}.$$

Since

$$\mu(T;0;\sigma) \to M_t\{g(\sigma,t)\} \text{ as } T \to \infty,$$

and since ϵ and η are independent of one another, while T is subjected only to the limitation $T > R(\eta)$, it follows, by letting first $\epsilon \to 0$, then $T \to \infty$, and finally $\eta \to 0$, that

$$M_t\{g(\sigma,t)\} \leqq \frac{\beta-\sigma}{\beta-\alpha}M_t\{g(\alpha,t)\} + \frac{\sigma-\alpha}{\beta-\alpha}M_t\{g(\beta,t)\}.$$

Since this inequality holds for every point σ of the interval $[\alpha,\beta]$, and since the latter can be replaced by any of its subintervals, the proof for the convexity of the function $M_t\{g(\sigma,t)\}$ of σ is complete.

QUEENS COLLEGE,
THE JOHNS HOPKINS UNIVERSITY.

NORMAL DISTRIBUTIONS AND THE LAW OF THE ITERATED LOGARITHM.*

By Philip Hartman.

1. Let $x_1(t), x_2(t), \cdots$ be a sequence of statistically independent functions defined on the interval $0 \leqq t \leqq 1$. Let the function $x_n(t)$ be normally distributed with mean 0 and variance b_n, that is, let

$$\text{meas}\,\{x_n < x\} = (2\pi b_n)^{-\frac{1}{2}} \int_{-\infty}^{x} \exp\,(-u^2/2b_n)\,du, \tag{1}$$

where meas $\{x_n < x\}$ denotes the Lebesgue measure of the t-set defined by the inequality $x_n(t) < x$. Finally, let B_n denote the variance of the function

$$s_n(t) = x_1(t) + \cdots + x_n(t), \tag{2}$$

so that,

$$B_n = b_1 + \cdots + b_n. \tag{3}$$

It is known that, if the numbers (3) are suitably restricted,[1] for example, if

$$B_n \to \infty, \text{ as } n \to \infty, \tag{4}$$

and

$$b_n = o(B_n/\log\log B_n), \text{ as } n \to \infty, \tag{5}$$

then

$$\overline{\lim_{n\to\infty}}\, s_n(t)/(B_n \log\log B_n)^{\frac{1}{2}} = \sqrt{2} \text{ for almost all } t. \tag{6}$$

The value of the quantity on the left of (6) under more general conditions has not appeared in the literature. The object of this note is to consider this upper limit when the only restriction imposed on the numbers B_n is (4).

It follows from (4) and from the "0 or 1" principle[2] that the expression on the left of (6) is a constant for almost all t. This constant may be described in terms of the B_n as follows:

---

* Received December 5, 1940.

[1] The literature does not seem to contain any explicit sufficient condition for (6) in the case of normally distributed functions. That the condition $0 < b_1 = b_2 = b_3 = \cdots$ is sufficient, may be deduced from the proof of Khintchine for continuous stochastic processes involving normal distributions (A. Khintchine, *Asymptotische Gesetze der Wahrsheintlichkeitsrechnung*, Berlin (1933), pp. 68-72). From the theorem in the paper, P. Hartman and A. Wintner, "On the law of the iterated logarithm," *American Journal of Mathematics*, vol. 63 (1941), pp. 169-176, it follows that (6) holds if $b_n = O(1)$ and $n = O(B_n)$. The sufficient conditions (4), (5) follow readily from a statement in J. Marcinkiewicz and A. Zygmund, "Remarque sur la loi du logarithme itéré," *Fundamenta Mathematicae*, vol. 29 (1937), see p. 222.

[2] Cf., e. g., A. Kolmogoroff, *Grundbegriffe der Wahrsheinlichkeitsrechnung*, Berlin (1933); in particular, see the appendix.

Let α be the number satisfying the requirements:

(i) *if r is any number greater than α and if $\{n_k\}$ is any sequence of integers for which there exists a constant $c > 1$ such that*

(7) $$B_{n_k} > cB_{n_{k-1}}, \qquad (k = 2, 3, \cdots),$$

then the series

(8) $$\Sigma \mid \log B_{n_k} \mid^{-r}$$

converges;

(ii) *if r is any number less than α, then there exists a sequence of integers $\{n_k\}$ for which there is a constant $c > 1$ satisfying (7) and for which the series (8) diverges. Then*

(9) $$\overline{\lim_{n\to\infty}}\, s_n(t)/(B_n \log\log B_n)^{\frac{1}{2}} = (2\alpha)^{\frac{1}{2}} \text{ for almost all } t.$$

It is clear from the definition of α that $0 \leqq \alpha \leqq 1$. Furthermore,[3] if

(5 bis) $$\overline{\lim_{n\to\infty}}\, b_n/B_n < 1,$$

then $\alpha = 1$; so that (6) holds even if (5) is replaced by (5 bis).

The question might arise whether or not (9) remains valid if the assumption concerning the normality of the distribution of $x_n(t)$ is dropped. This question must be answered in the negative; in fact,[4] (9) need not hold even in the case where each $x_n(t)$ assumes only two values, $\pm \sqrt{b_n}$.

In the next section, a simplification of the definition of the number α will be obtained. In Section 3, it will be shown that (i) implies the relation (9) with "$=$" replaced by "$\leqq$". Finally, the relation (9) with "$=$" replaced by "$\geqq$" will be deduced from (ii) in Section 4. In this last section, the result of Section 3 is used, but this is unnecessary. The methods used in Sections 3 and 4 will be similar to the procedure occurring in Kolmogoroff's paper[5] on the law of the iterated logarithm.

2. The above definition of α involves all sequences of integers $\{n_k\}$ for which there is a constant $c > 1$ satisfying (7); actually, it is sufficient to consider one such sequence, provided that an additional condition is imposed.

---

[3] It is clear that in determining α, it is only necessary to consider sequences $\{n_k\}$ for which (7) and $B_{n_{k-1}} \leqq cB_{n_{k-1}}$ are satisfied. The condition (5 bis) implies that for such sequences $\mid \log B_{n_k} \mid / k$ lies between two positive constants. See Section 2, also.

[4] It is easy to see that (9) is false for the example constructed by J. Marcinkiewicz and A. Zygmund, *loc. cit.*[1]. Their inequalities also show that if $x_n(t)$ takes only two values, $\pm\sqrt{b_n}$, then the upper limit occurring in (9) does not exceed $(2\alpha)^{\frac{1}{2}}$ for almost all t.

[5] A. Kolmogoroff, "Über das Gesetz des iterierten Logarithmus," *Mathematische Annalen*, vol. 101 (1929), pp. 126-135.

Let $C > 1$ be a fixed number and let $\{m_k\}$ be any fixed sequence of integers such that

$$(10_1) \qquad B_{m_k-1} \leqq CB_{m_{k-1}}; \qquad\qquad (10_2) \qquad B_{m_k} > CB_{m_{k-1}}.$$

Let β be the greatest lower bound of all numbers r for which the series

$$(11) \qquad \Sigma \mid \log B_{m_k} \mid^{-r}$$

converges. Then

$$(12) \qquad \beta = \alpha.$$

In order to prove (12), note that if r is any number greater than α, then, in view of (10_2) and (i), the series (11) converges. Hence, $r \geqq \beta$. It follows that $\alpha \geqq \beta$. On the other hand, if r is any number less than α, let $\{n_k\}$ and c be a sequence of integers and a constant, respectively, whose existence is assured by (ii). Finally, let j be a positive integer such that $c^j > C$. The divergence of the series (8) implies that there exists at least one integer h, such that $0 \leqq h < j$ and

$$(13) \qquad \sum_{k=1}^{\infty} \mid \log B_{j_k} \mid^{-r} = \infty; \qquad j_k = n_{kj+h}, \qquad (k = 1, 2, \cdots).$$

We also have

$$(14) \qquad B_{j_k} > c^j B_{j_{k-1}} > CB_{j_{k-1}}, \qquad (k = 2, 3, \cdots).$$

Without violating (13) or (14), it may be supposed that $B_{j_1} \geqq B_{m_1}$; in which case it follows from (10_1), (10_2) and (14), that

$$B_{j_k} \geqq B_{m_k}, \qquad (k = 1, 2, \cdots).$$

Consequently, (13) implies the divergence of the series (11). Hence, $r \leqq \beta$. It follows that $\alpha \leqq \beta$. This completes the proof of (12).

3. Let $\delta > 0$ be arbitrarily fixed. Define a sequence of integers $\{n_k\}$ by induction, as follows: Choose n_1 such that $B_n > e$ if $n \geqq n_1$; and if n_{k-1} has been defined, let n_k be determined by the inequalities

$$(15_1) \qquad B_{n_k-1} \leqq (1 + \delta) B_{n_{k-1}}; \qquad (15_2) \qquad B_{n_k} > (1 + \delta) B_{n_{k-1}}.$$

If t, $0 \leqq t \leqq 1$, is such that the function (2) satisfies

$$s_n(t) > [2(1 + 3\delta)\alpha B_n \log\log B_n]^{\frac{1}{2}},$$

for at least one integer n in the interval $n_{k-1} \leqq n < n_k$, then

$$(16) \qquad S_{n_k-1} > [2(1 + 3\delta)\alpha B_{n_{k-1}} \log\log B_{n_{k-1}}]^{\frac{1}{2}},$$

where

$$S_n(t) = \max [s_1(t), \cdots, s_n(t)].$$

On denoting by V_k the measure of the t-set defined by the inequality (16), it follows from a well-known lemma[6] of Kolmogoroff that

$$V_k \leqq 2 \text{ meas } \{s_{n_{k-1}} > [2(1+3\delta)\alpha B_{n_{k-1}} \log\log B_{n_{k-1}}]^{\frac{1}{2}} - (2B_{n_k-1})^{\frac{1}{2}}\}.$$

Hence, for sufficiently large k,

$$V_k \leqq 2 \text{ meas } \{s_{n_k-1} > [2(1+2\delta)\alpha B_{n_{k-1}} \log\log B_{n_{k-1}}]^{\frac{1}{2}}\},$$

since (15_1) and (4) imply

$$B_{n_k-1}/B_{n_{k-1}} \log\log B_{n_{k-1}} \to 0, \text{ as } k \to \infty.$$

From the statistical independence of $x_1(t), \cdots, x_n(t)$ and from (1), (2), (3),

$$\text{meas } \{s_n(t) > x\} = (2\pi B_n)^{-\frac{1}{2}} \int_x^\infty \exp(-u^2/2B_n)\,du;$$

so that,[7] for sufficiently large k,

$$V_k < \exp[-(1+2\delta)\alpha B_{n_{k-1}} \log\log B_{n_{k-1}}/B_{n_k-1}].$$

Hence, by (15_1),

(17) $\qquad V_k < (\log B_{n_{k-1}})^{-r}, \quad \text{where} \quad r = (1+2\delta)\alpha/(1+\delta) > \alpha.$

Since (7) is implied by (15_2) for $c = 1 + \delta > 1$, it follows from (17) and from the property (i) of α that

$$\sum_{k=1}^{\infty} V_k < \infty.$$

Consequently, for any $\delta > 0$ and for almost all t, there exists an integer $n = n_0(t)$ such that

(18) $\qquad s_n(t) < [2(1+\delta)\alpha B_n \log\log B_n]^{\frac{1}{2}} \text{ for } n > n_0(t).$

This is equivalent to the statement that

$$\overline{\lim_{n\to\infty}}\, s_n(t)/(B_n \log\log B_n)^{\frac{1}{2}} \leqq (2\alpha)^{\frac{1}{2}} \text{ for almost all } t.$$

4. In order to prove the reverse inequality, let $\delta > 0$ again be arbitrarily fixed. Since $\alpha(1-\delta) < \alpha$, it follows from the property (ii) of α that there exists a sequence of integers $\{n_k\}$ for which there is a constant $c > 1$ satisfying (7) and such that

(19) $\qquad \sum_{k=1}^{\infty} |\log B_{n_k}|^{-r} = \infty \text{ if } r = \alpha(1-\delta).$

---

[6] *Loc. cit.*[5], Hilfssätz 5, p. 131.

[7] This is a consequence of the asymptotic relation

$$\int_x^\infty \exp(-u^2/2)\,du \backsim \exp(-x^2/2)/x, \text{ as } x \to +\infty.$$

It may be supposed that the number c in (7) is arbitrarily large (but fixed), for $c(>1)$ may always be replaced by c^2 by replacing the sequence $\{n_k\}$ by one of the two sequences $\{n_{2k}\}$, $\{n_{2k+1}\}$; cf. Section 2.

Let U_k denote the measure of the t-set defined by

$$(20)\qquad s_{n_k}(t) - s_{n_{k-1}}(t) > [2(1-2\delta)\alpha\beta_k \log\log \beta_k]^{\frac{1}{2}},$$

where

$$(21)\qquad \beta_k = B_{n_k} - B_{n_{k-1}}.$$

In virtue of the fact that $s_{n_k}(t) - s_{n_{k-1}}(t)$ is normally distributed with the mean 0 and the variance β_k, it is easy to see that,[7] for sufficiently large k,

$$U_k > (\log \beta_k)^{-r}, \qquad r = \alpha(1-\delta).$$

Since $\beta_k < B_{n_k}$, it follows from (19), that

$$(22)\qquad \sum_{k=1}^{\infty} U_k = \infty.$$

Now, if t, $0 \leqq t \leqq 1$, is such that (20) holds for only a finite number of values of k, then, since the functions $s_{n_2} - s_{n_1}$, $s_{n_3} - s_{n_2}, \cdots$ are statistically independent, t belongs to a set whose measure does not exceed

$$\sum_{n=2}^{\infty} \prod_{k=n}^{\infty} (1 - U_k).$$

But this last expression is equal to 0 by (22). Consequently, (20) holds for infinitely many values of k for every t belonging to a set of measure 1.

Let t be such that (18) holds; so that, in particular,

$$s_{n_{k-1}}(t) < [2(1+\delta)\alpha B_{n_{k-1}} \log\log B_{n_{k-1}}]^{\frac{1}{2}},$$

for every sufficiently large k. If, in addition, (20) holds, then

$$(23)\quad s_{n_k}(t) > [2(1-2\delta)\alpha\beta_k \log\log \beta_k]^{\frac{1}{2}} - [2(1+\delta)\alpha B_{n_{k-1}} \log\log B_{n_{k-1}}]^{\frac{1}{2}}.$$

But, by (7),

$$(24)\qquad B_{n_{k-1}} < B_{n_k}/c;$$

so that, by (21),

$$(25)\qquad \beta_k > B_{n_k}(1 - 1/c).$$

Hence, if $c = c(\delta)$ is sufficiently large, (23), (24) and (25) give

$$(26)\qquad s_{n_k}(t) > [2(1-3\delta)B_{n_k} \log\log B_{n_k}]^{\frac{1}{2}}.$$

It follows that, for every t belonging to a set of measure 1, (26) holds for an infinity of values of k. This completes the proof of (9).

QUEENS COLLEGE.

ON SOME PARTITION FUNCTIONS.*

By Mary Haberzetle.

1. Introduction. Let p and q be distinct primes satisfying the condition that the product $(p-1)(q-1)$ be divisible by 24. If μ is an integer defined by $24\mu-(p-1)(q-1)=0$, then for every integer $m>\mu$ we shall obtain a formula for the number of partitions of m into parts none of which are multiples of p or q. This formula involves the number of partitions of $1, 2, \cdots, \mu-1$ into summands of this type.

The method we shall use is essentially that used by Rademacher [1] in finding the Fourier coefficients of the modular invariant $J(\tau)$. A related paper is that of Ivan Niven,[2] in which a formula for the number of partitions of an integer into parts, none of which are multiples of 2 or 3, is obtained. This result is obviously not included in the problem set forth above.

The number of partitions of an integer m into parts, none of which are multiples of p or q, appears as the coefficient of x^m in the expansion of

$$F(x)=\frac{f(x)\,f(x^{pq})}{f(x^p)\,f(x^q)},$$

where

$$f(x)=\frac{1}{(1-x)(1-x^2)(1-x^3)\cdots}.$$

Denote the number of such partitions by a_m. Applying Cauchy's theorem to $F(x)$, we obtain

$$a_m=\frac{1}{2\pi i}\int_C\frac{F(x)}{x^{m+1}}\,dx,$$

where C is a closed curve surrounding $x=0$ and inside $|x|=1$. Take C to be the circle defined by $|x|=\exp(-2\pi N^{-2})$, where N is an arbitrary positive integer. Consider the Farey series of order N, and divide the circle C into Farey arcs $\xi_{h,k}$. In this way we obtain

$$(1.1)\qquad a_m=\sum_{\substack{h,k\\0\leqq h<k\leqq N}}{}'\;\frac{1}{2\pi i}\int_{\xi_{h,k}}\frac{F(x)}{x^{m+1}}\,dx.$$

---

* Received August 3, 1940.

[1] "The Fourier coefficients of the modular invariant $J(\tau)$," *American Journal of Mathematics*, vol. 60 (1938), pp. 501-512.

[2] "On a certain partition function," *ibid.*, vol. 62 (1940), pp. 353-364.

Here and throughout this paper Σ' denotes a summation in which h takes only integral values prime to k.

On the Farey arc $\xi_{h,k}$

$$x = \exp\left(-2\pi N^{-2} + \frac{2\pi ih}{k} + 2\pi i\phi\right),$$

where $-\theta'_{h,k} \leqq \phi \leqq \theta''_{h,k}$. For later application we shall now exhibit $\theta'_{h,k}$ and $\theta''_{h,k}$ as determined by h and k.

If $\frac{h_1}{k_1}, \frac{h_2}{k_2}$ are the left and right neighbors of $\frac{h}{k}$ in the Farey series of order N, then

$$(1.2) \qquad \frac{h_1}{k_1} < \frac{h}{k} < \frac{h_2}{k_2}, \qquad k, k_1, k_2 \leqq N,$$

and

$$hk_1 - h_1k = 1, \qquad h_2k - hk_2 = 1.$$

Hence,

$$(1.3) \qquad hk_1 \equiv 1 \pmod{k}, \qquad hk_2 \equiv -1 \pmod{k}.$$

Later we shall use an integer h' determined by $hh' \equiv -1 \pmod{k}$. Then, from (1.3),

$$(1.4) \qquad k_1 \equiv -h' \pmod{k}, \qquad k_2 \equiv h' \pmod{k}.$$

The Farey segment around $\frac{h}{k}$ is bounded by the mediants,

$$\frac{h_1 + h}{k_1 + k}, \qquad \frac{h_2 + h}{k_2 + k}.$$

Since these do not belong to the Farey series of order N we have

$$k_1 + k > N, \qquad k_2 + k > N.$$

These conditions, along with (1.2), restrict k_1 and k_2 to the intervals,

$$(1.5) \qquad N - k < k_1 \leqq N, \qquad N - k < k_2 \leqq N.$$

The relations (1.4) and (1.5) determine k_1 and k_2 uniquely as functions of h and k, and it is known from the theory of Farey series that

$$\theta'_{h,k} = \frac{1}{k(k_1 + k)}, \qquad \theta''_{h,k} = \frac{1}{k(k_2 + k)}.$$

Introducing the new variable ϕ in (1.1), we get

$$a_m = \sum_{\substack{h,k \\ 0 \leqq h < k \leqq N}}' \int_{-\theta'_{h,k}}^{\theta''_{h,k}} F\left[\exp\left(\frac{2\pi ih}{k} - 2\pi(N^{-2} - i\phi)\right)\right]$$
$$\times \exp\left[-m\left(-2\pi N^{-2} + \frac{2\pi ih}{k} + 2\pi i\phi\right)\right] d\phi,$$

or, if $w = N^{-2} - i\phi$ and $z = kw$, then

$$(1.6) \quad a_m = \sum_{\substack{h,k \\ 0 \leq h < k \leq N}}' \exp\left(-\frac{2\pi i h m}{k}\right) \int_{-\theta'_{h,k}}^{\theta''_{h,k}} F\left[\exp\left(\frac{2\pi i h}{k} - \frac{2\pi z}{k}\right)\right] \exp(2\pi m w)\, d\phi.$$

Next we shall apply to $F(x)$ in the integrand of (1.6) a transformation derived from the theory of modular functions. In the next section we shall display the formulas needed for this transformation.

2. Transformation formulas for $F(x)$. From the theory of modular functions we obtain the identity,

$$f\left[\exp\left(\frac{2\pi i h}{k} - \frac{2\pi z}{k}\right)\right] = \omega_{h,k} z^{\frac{1}{2}} \exp\left(\frac{\pi}{12kz} - \frac{\pi z}{12k}\right) f\left[\exp\left(\frac{2\pi i h'}{k} - \frac{2\pi}{kz}\right)\right],$$

valid for $\Re(z) > 0$ with $\omega_{h,k}$ defined by

$$\omega_{h,k} = \left(\frac{-h}{h}\right) \exp\left\{-\pi i\left[\frac{1}{4}(k-1) + \frac{1}{12}\left(k - \frac{1}{k}\right)(2h + h'h^2 - h')\right]\right\}$$

for odd k and by

$$\omega_{h,k} = \left(\frac{-k}{k}\right) \exp\left\{-\pi i\left[\frac{1}{4}(2 - h - hk) + \frac{1}{12}\left(k - \frac{1}{k}\right)(2h + h'h^2 - h')\right]\right\}$$

for odd h. The notation $\left(\frac{a}{b}\right)$ designates the Legendre-Jacobi symbol, and h' is any solution of the congruence, $hh' \equiv -1 \pmod{k}$.

Applying this to $F(x)$ we obtain

$$(2.1) \quad F\left[\exp\left(\frac{2\pi i h}{k} - \frac{2\pi z}{k}\right)\right] = \Omega_{h,k}\psi_k(z)\; F\left[\exp\left(\frac{2\pi i h'}{k} - \frac{2\pi}{kz}\right)\right],$$

$$(2.2) \quad F\left[\exp\left(\frac{2\pi i h}{k} - \frac{2\pi z}{k}\right)\right] = \Omega_{h,k}\psi_k(z)\; F^{-1}\left[\exp\left(\frac{2\pi i h'}{qk} - \frac{2\pi}{qkz}\right)\right],$$

$$(2.3) \quad F\left[\exp\left(\frac{2\pi i h}{k} - \frac{2\pi z}{k}\right)\right] = \Omega_{h,k}\psi_k(z)\; F^{-1}\left[\exp\left(\frac{2\pi i h'}{pk} - \frac{2\pi}{pkz}\right)\right],$$

$$(2.4) \quad F\left[\exp\left(\frac{2\pi i h}{k} - \frac{2\pi z}{k}\right)\right] = \Omega_{h,k}\psi_k(z)\; F\left[\exp\left(\frac{2\pi i h'}{pqk} - \frac{2\pi}{pqkz}\right)\right],$$

according as $(k, pq) = pq, p, q$, or 1, respectively. In all four cases $hh' \equiv -1 \pmod{k}$, and in the second, third, and fourth cases h' is chosen to be divisible by q, p, and pq, respectively.

The symbols, $\Omega_{h,k}$ and $\psi_k(z)$, denote the following:

$$(2.5) \quad \Omega_{h,k} = \frac{\omega_{h,k}\omega_{h,k/pq}}{\omega_{h,k/p}\omega_{h,k/q}}, \quad \psi_k(z) = \exp\left[\frac{\pi}{12k}(p-1)(q-1)\left(\frac{1}{z} - z\right)\right],$$
$$\text{for } (k, pq) = pq,$$

$$(2.6) \quad \Omega_{h,k} = \frac{\omega_{h,k}\omega_{qh,k/p}}{\omega_{h,k/p}\omega_{qh,k}}, \quad \psi_k(z) = \exp\left[-\frac{\pi}{12k}(p-1)(q-1)\left(\frac{1}{qz} + z\right)\right],$$
$$\text{for } (k, pq) = p,$$

$$(2.7)\quad \Omega_{h,k}=\frac{\omega_{h,k}\omega_{ph,k/q}}{\omega_{h,k/q}\omega_{ph,k}},\quad \psi_k(z)=\exp\left[-\frac{\pi}{12k}(p-1)(q-1)\left(\frac{1}{pz}+z\right)\right],$$
$$\text{for } (k,pq)=q,$$

$$(2.8)\quad \Omega_{h,k}=\frac{\omega_{h,k}\omega_{pqh,k}}{\omega_{ph,k}\omega_{qh,k}},\quad \psi_k(z)=\exp\left[\frac{\pi}{12k}(p-1)(q-1)\left(\frac{1}{pqz}-z\right)\right],$$
$$\text{for } (k,pq)=1.$$

3. Estimate for $a_m^{(pq)}$. In order to apply the transformation formulas above to (1.6) we decompose a_m into parts $a_m^{(pq)}$, $a_m^{(p)}$, $a_m^{(q)}$, $a_m^{(1)}$, according as $(k,pq)=pq$, p, q, or 1, respectively. We shall estimate each in turn.

From (1.6), (2.1), (2.5), and $F(x)=\sum_{n=0}^{\infty}a_nx^n$ we obtain

$$(3.1)\quad \begin{aligned} a_m^{(pq)} = &\sum_{\substack{k=1\\(k,pq)=pq}}^{N}\ \sum_{n=0}^{\mu-1}a_n\sum_{h\bmod k}{}'\exp\left[-\frac{2\pi i}{k}(mh-nh')\right]\Omega_{h,k}\\ &\times\int_{-\theta'_{h,k}}^{\theta''_{h,k}}\exp\left[-\frac{2\pi}{k^2w}(n-\mu)+2\pi w(m-\mu)\right]d\phi\\ &+\sum_{\substack{k=1\\(k,pq)=pq}}^{N}\ \sum_{n=\mu}^{\infty}a_n\sum_{h\bmod k}{}'\exp\left[-\frac{2\pi i}{k}(mh-nh')\right]\Omega_{h,k}\\ &\times\int_{-\theta'_{h,k}}^{\theta''_{h,k}}\exp\left[-\frac{2\pi}{k^2w}(n-\mu)+2\pi w(m-\mu)\right]d\phi=I_1+I_2. \end{aligned}$$

The integer μ is defined by $24\mu-(p-1)(q-1)=0$. We shall assume throughout this section that $(k,pq)=pq$ without hereafter noting it explicitly under Σ. Note also that h is any solution of $hh'\equiv-1\pmod k$.

We shall first find an estimate for I_2. Using

$$(3.2)\quad -\theta'_{h,k}=-\frac{1}{k(k_1+k)}\leqq-\frac{1}{k(N+k)}<\frac{1}{k(N+k)}\leqq\frac{1}{k(k_2+k)}=\theta''_{h,k}$$

we may write

$$\begin{aligned} I_2=&\sum_{k=1}^{N}\sum_{n=\mu}^{\infty}a_n\int_{-[1/k(N+k)]}^{1/k(N+k)}\exp\left[-\frac{2\pi}{k^2w}(n-\mu)+2\pi w(m-\mu)\right]d\phi\\ &\times\sum_{h\bmod k}{}'\exp\left[-\frac{2\pi i}{k}(mh-nh')\right]\Omega_{h,k}\\ &+\sum_{k=1}^{N}\sum_{n=\mu}^{\infty}a_n\sum_{h\bmod k}{}'\exp\left[-\frac{2\pi i}{k}(mh-nh')\right]\Omega_{h,k}\sum_{l=k_1+k}^{N+k-1}\int_{-[1/kl]}^{-[1/k(l+1)]}\\ &+\sum_{k=1}^{N}\sum_{n=\mu}^{\infty}a_n\sum_{h\bmod k}{}'\exp\left[-\frac{2\pi i}{k}(mh-nh')\right]\Omega_{h,k}\sum_{l=k_2+k}^{N+k-1}\int_{1/k(l+1)}^{1/kl}=S_1+S_2+S_3, \end{aligned}$$

where the integrand is the same in all three expressions.

The inner sum in S_1 is equal to

$$\sum_{h \bmod k}{}' \exp \{-\frac{2\pi i}{k}[h(m-\mu)+h'(-n+\mu)]\}. \tag{3.3}$$

This is a Kloosterman sum modulo k, and from the results of Kloosterman,[3] Estermann,[4] Salie,[5] and Davenport,[6] we have the following estimate for it:

$$O\{k^{(2/3)+\epsilon}(m-\mu, k)^{1/3}\}.$$

This may be replaced by

$$O\{k^{(2/3)+\epsilon}m^{1/3}\}$$

since $m > \mu$.

For $\mu < n \leqq 2\mu - 1$,

$$\Re\left[\frac{2\pi}{k^2 w}(n-\mu)\right] \geqq \Re\left(\frac{2\pi}{k^2 w}\right) = \frac{2\pi N^{-2}}{k^2(N^{-4}+\phi^2)} \geqq \pi. \tag{3.4}$$

For $n \geqq 2\mu$,

$$\Re\left[\frac{2\pi}{k^2 w}(n-\mu)\right] \geqq \Re\left(\frac{\pi n}{k^2 w}\right) = \frac{\pi n N^{-2}}{k^2(N^{-4}+\phi^2)} \geqq \frac{\pi n}{2}. \tag{3.5}$$

Hence,

$$\begin{aligned} S_1 &= O\left\{\sum_{k=1}^{N} \frac{2}{kN} \exp(2\pi m N^{-2}) k^{(2/3)+\epsilon} n^{1/3} \left[a_\mu + \sum_{n=\mu+1}^{2\mu-1} a_n \exp(-\pi) \right.\right. \\ &\qquad \left.\left. + \sum_{n=2\mu}^{\infty} a_n \exp\left(-\frac{\pi n}{2}\right)\right]\right\} \\ &= O\left\{\frac{1}{N}\sum_{k=1}^{N} k^{(-1/3)+\epsilon} \exp(2\pi m N^{-2}) m^{1/3}\right\} \\ &= O\{m^{1/3} N^{(-1/3)+\epsilon} \exp(2\pi m N^{-2})\}. \end{aligned}$$

Because of the similarity of S_2 and S_3 we shall treat only the latter. Interchanging the summation with respect to h and l we obtain

$$\begin{aligned} S_3 = \sum_{k=1}^{N}\sum_{n=\mu}^{\infty} a_n \sum_{l=N+1}^{N+k-1} \int_{1/k(l+1)}^{1/kl} \exp\left[-\frac{2\pi}{k^2 w}(n-\mu) + 2\pi w(m-\mu)\right] d\phi \\ \times \sum_{\substack{h \bmod k \\ N < k+k_2 \leqq l}}{}' \exp\left[-\frac{2\pi i}{k}(mh - nh')\right]\Omega_{h,k}. \end{aligned}$$

The inner sum is like (3.3) but has the added condition that $N < k + k_2 \leqq l$. This restriction on k_2 implies, in consequence of (1.4), a restriction on h'

---

[3] "Asymptotische Formeln für die Fourierkoeffizienten ganzer Modulformen," *Abhandlungen Hamburg. Math. Seminar*, vol. 5 (1927), pp. 337-352.

[4] "Vereinfachter Beweis eines Satzes von Kloosterman," *ibid.*, vol. 7 (1929), pp. 82-98.

[5] "Zur Abschätzung der Fourierkoeffizienten ganzer Modulformen," *Mathematische Zeitschrift*, vol. 36 (1933), pp. 263-278.

[6] "On certain exponential sums," *Journal f. d. reine u. angew. Mathematik*, vol. 169 (1933), pp. 158-176.

to an interval modulo k equivalent to one or two intervals in the range $0 \leqq h' < k$. Hence our present sum, corresponding to (3.3), is an incomplete Kloosterman sum modulo k, for which we have the estimate $O\{k^{(2/3)+\epsilon}m^{1/3}\}$. The relations (3.4) and (3.5) are valid for the case of S_3, and we obtain

$$\begin{aligned} S_3 &= O\left\{\sum_{k=1}^{N} \exp\,(2\pi m N^{-2}) k^{(2/3)+\epsilon} m^{1/3} \sum_{l=N+1}^{N+k-1} \left[\frac{1}{kl} - \frac{1}{k(l+1)}\right]\right. \\ &\qquad \left. \times \left[a_\mu + \sum_{n=\mu+1}^{2\mu-1} a_n \exp\,(-\pi) + \sum_{n=2\mu}^{\infty} a_n \exp\left(-\frac{\pi n}{2}\right)\right]\right\} \\ &= O\left\{\frac{1}{N}\sum_{k=1}^{N} k^{(-1/3)+\epsilon} \exp\,(2\pi m N^{-2}) m^{1/3}\right\} \\ &= O\{m^{1/3} N^{(-1/3)+\epsilon} \exp\,(2\pi m N^{-2})\}. \end{aligned}$$

A like result holds for S_2. Therefore,

$$I_2 = O\{m^{1/3} N^{(-1/3)+\epsilon} \exp\,(2\pi m N^{-2})\}. \tag{3.6}$$

From (3.1) and (3.2) we may write

$$\begin{aligned} I_1 &= \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n \int_{-[1/k(N+k)]}^{1/k(N+k)} \exp\left[-\frac{2\pi}{k^2 w}(n-\mu) + 2\pi w(m-\mu)\right] d\phi \\ &\qquad \times \sum_{h \bmod k}{}' \exp\left[-\frac{2\pi i}{k}(mh - nh')\right]\Omega_{h,k} \\ &\quad + \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n \sum_{h \bmod k}{}' \exp\left[-\frac{2\pi i}{k}(mh - nh')\right]\Omega_{h,k} \sum_{l=k_1+k}^{N+k-1} \int_{-[1/kl]}^{-[1/k(l+1)]} \\ &\quad + \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n \sum_{h \bmod k}{}' \exp\left[-\frac{2\pi i}{k}(mh - nh')\right]\Omega_{h,k} \sum_{l=k_2+k}^{N+k-1} \int_{1/k(l+1)}^{1/kl} \\ &= Q_0 + Q_1 + Q_2, \end{aligned} \tag{3.7}$$

where all three integrals have the same integrand.

The essential difference between I_1 and I_2 is that in I_1 the coefficient of $\frac{1}{w}$ in the exponent of the integrand is positive while in I_2 it is negative or zero. Define

$$B_{k,n}(m) = \sum_{h \bmod k}{}' \exp\left[-\frac{2\pi i}{k}(mh - nh')\right]\Omega_{h,k}, \tag{3.8}$$

$\Omega_{h,k}$ being given by (2.5). Consider in the complex w-plane the closed rectangular path R with vertices

$$\pm N^{-2} \pm \frac{i}{k(N+k)},$$

and take R as surrounding 0 in the positive sense. Then we may write

$$Q_0 = 2\pi \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m) \frac{1}{2\pi i} \int_{R} \exp\left[-\frac{2\pi}{k^2 w}(n-\mu) + 2\pi w(m-\mu)\right] dw$$

$$-\frac{1}{i} \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m) \left\{ \int_{N^{-2}+i/k(N+k)}^{-N^{-2}+i/k(N+k)} + \int_{-N^{-2}+i/k(N+k)}^{-N^{-2}-i/k(N+k)} + \int_{-N^{-2}-i/k(N+k)}^{N^{-2}-i/k(N+k)} \right\}$$

$$= 2\pi \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m)\, L_{k,n}(m)$$

$$-\frac{1}{i} \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m) \{J_1 + J_2 + J_3\},$$

where the integrand is the same in all four integrals.

The integral

$$L_{k,n}(m) = \frac{1}{2\pi i} \int_{R} \exp\left[-\frac{2\pi}{k^2 w}(n-\mu) + 2\pi w(m-\mu)\right] dw$$

$$= \frac{1}{2\pi i} \int_{R} \sum_{\alpha=0}^{\infty} \frac{\left[\frac{2\pi}{k^2 w}(\mu-n)\right]^{\alpha}}{\alpha!} \sum_{\beta=0}^{\infty} \frac{[2\pi w(m-\mu)]^{\beta}}{\beta!}\, dw$$

$$= \frac{1}{k} \sqrt{\frac{\mu-n}{m-\mu}} \sum_{\alpha=0}^{\infty} \frac{\left(\frac{2\pi}{k}\sqrt{(\mu-n)(m-\mu)}\right)^{2\alpha+1}}{\alpha!\,(\alpha+1)!},$$

or

$$L_{k,n}(m) = \frac{1}{k} \sqrt{\frac{\mu-n}{m-\mu}}\, I_1\left(\frac{4\pi}{k}\sqrt{(\mu-n)(m-\mu)}\right), \tag{3.9}$$

where $I_1(z)$ is the Bessel function of first order with pure imaginary argument.[7]

Along the paths of integration for J_1 and J_3

$$w = u \pm \frac{i}{k(N+k)}, \qquad -N^{-2} \leqq u \leqq N^{-2}.$$

Hence we have

$$\Re(w) = u \leqq N^{-2},\ \Re\left(\frac{1}{w}\right) = \frac{u}{u^2 + k^{-2}(N+k)^{-2}} < N^{-2}k^2(N+k)^2 \leqq 4k^2,$$

so that the absolute value of the integrand in J_1 and J_3 is $< \exp(8\pi\mu + 2\pi m N^{-2})$. Therefore,

$$\begin{vmatrix} J_1 \\ J_3 \end{vmatrix} < 2N^{-2} \exp(8\pi\mu + 2\pi m N^{-2}).$$

In J_2, $w = -N^{-2} + iv$, where $-\frac{1}{k(N+k)} \leqq v \leqq \frac{1}{k(N+k)}$. Hence, $\Re(w) = -N^{-2} < 0$, $\Re\left(\frac{1}{w}\right) = \frac{-N^{-2}}{N^{-4}+v^2} < 0$, so that the absolute value of

[7] G. N. Watson, *Theory of Bessel Functions*, Cambridge (1922), p. 77.

the integrand is < 1. Therefore

$$| J_2 | \leqq \frac{2}{k(N+k)} < 2k^{-1} N^{-1}.$$

From (3.8), (3.3), and the discussion following (3.3) we have $B_{k,n}(m) = O\{k^{(2/3)+\epsilon} m^{1/3}\}$. Collecting our results, we obtain

$$Q_0 = 2\pi \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m) L_{k,n}(m) + O\{\exp(2\pi m N^{-2}) \frac{1}{N} \sum_{k=1}^{N} k^{((-1/3)+\epsilon} m^{1/3}\},$$

or

$$Q_0 = 2\pi \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m) L_{k,n}(m) + O\{m^{1/3} N^{(-1/3)+\epsilon} \exp(2\pi m N^{-2})\}. \tag{3.10}$$

Interchanging the order of summation on l and h we get from (3.7)

$$Q_1 = \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n \sum_{l=N+1}^{N+k-1} \int_{-[1/kl]}^{-[1/k(l+1)]} \exp[-\frac{2\pi}{k^2 w}(n-\mu) + 2\pi w(m-\mu)] d\phi$$

$$\times \sum_{\substack{h \bmod k \\ N < k+k_1 \leqq l}}' \exp[-\frac{2\pi i}{k}(mh - nh')] \Omega_{h,k}$$

For the inner sum we again have the estimate $O\{k^{(2/3)+\epsilon} m^{1/3}\}$. Since

$$\Re\left[\frac{2\pi}{k^2 w}(\mu - n)\right] \leqq \frac{2\pi(\mu - n)}{k^2 N^2 [N^{-4} + k^{-2}(N+k)^{-2}]} < 8\pi\mu$$

and

$$\Re[2\pi w(m-\mu)] < 2\pi m N^{-2},$$

we have

$$Q_1 = O\left\{ \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n \sum_{l=N+1}^{N+k-1} \left[\frac{1}{kl} - \frac{1}{k(l+1)}\right] \exp(8\pi\mu + 2\pi m N^{-2}) k^{(2/3)+\epsilon} m^{1/3} \right\},$$

or

$$Q_1 = O\{m^{1/3} N^{(-1/3)+\epsilon} \exp(2\pi m N^{-2})\}. \tag{3.11}$$

Because of the similarity between Q_1 and Q_2 we can deduce the same estimate for Q_2. Combining our estimates for Q_0, Q_1, and Q_2 from (3.10) and (3.11) we obtain

$$I_1 = 2\pi \sum_{k=1}^{N} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m) L_{k,n}(m) + O\{m^{1/3} N^{(-1/3)+\epsilon} \exp(2\pi m N^{-2})\}. \tag{3.12}$$

In consequence of (3.1), (3.6), and (3.12) we now have

$$a_m^{(pq)} = 2\pi \sum_{\substack{k=1 \\ (k,pq)=pq}}^{N} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m) L_{k,n}(m) \tag{3.13}$$

$$+ O\{m^{1/3} N^{(-1/3)+\epsilon} \exp(2\pi m N^{-2})\},$$

where $B_{k,n}(m)$ is given by (3.8) and $L_{k,n}(m)$ by (3.9).

4. Estimates for $a_m^{(p)}$ and $a_m^{(q)}$. From (1.6), (2.2), (2.6) and $F^{-1}(x) = \sum_{n=0}^{\infty} b_n x^n$ we obtain

$$a_m^{(p)} = \sum_{\substack{k=1\\(k,pq)=p}}^{N} \sum_{n=0}^{\infty} b_n \sum_{h \bmod k}{}' \exp\left[-\frac{2\pi i}{k}\left(mh - \frac{nh'}{q}\right)\right] \Omega_{h,k}$$
$$\times \int_{-\theta'_{h,k}}^{\theta''_{h,k}} \exp\left[-\frac{2\pi}{qk^2 w}(n+\mu) + 2\pi w(m-\mu)\right] d\phi.$$

Throughout this section we shall assume $(k, pq) = p$ without noting it explicitly hereafter. It should also be remembered that we now use a solution h' of $hh' \equiv -1 \pmod{k}$ which is divisible by q. Making use of (3.2) we may write

$$a_m^{(p)} = \sum_{k=1}^{N} \sum_{n=0}^{\infty} b_n \int_{-[1/k(N+k)]}^{1/k(N+k)} \exp\left[-\frac{2\pi}{k^2 qw}(n+\mu) + 2\pi w(m-\mu)\right] d\phi$$
$$\times \sum_{h \bmod k}{}' \exp\left[-\frac{2\pi i}{k}\left(mh - \frac{nh'}{q}\right)\right] \Omega_{h,k}$$
$$(4.1) \qquad + \sum_{k=1}^{N} \sum_{n=0}^{\infty} b_n \sum_{h \bmod k}{}' \exp\left[-\frac{2\pi i}{k}\left(mh - \frac{nh'}{q}\right)\right] \Omega_{h,k} \sum_{l=k_1+k}^{N+k-1} \int_{-[1/kl]}^{-[1/k(l+1)]}$$
$$+ \sum_{k=1}^{N} \sum_{n=0}^{\infty} b_n \sum_{h \bmod k}{}' \exp\left[-\frac{2\pi i}{k}\left(mh - \frac{nh'}{q}\right)\right] \Omega_{h,k} \sum_{l=k_2+k}^{N+k-1} \int_{1/k(l+1)}^{1/kl}$$
$$= R_1 + R_2 + R_3.$$

All three integrals have the same integrand. The coefficient of $\frac{1}{w}$ is now always negative, and we proceed as we did with I_2 for $a_m^{(pq)}$.

The inner sum in R_1 is equal to

$$(4.2) \qquad \pm \sum_{h \bmod k}{}' \exp\left\{-\frac{2\pi i}{k}\left[h(m-\mu) + \frac{h'}{q}(-n-\mu)\right]\right\}.$$

Determine a by $aq \equiv -1 \pmod{k}$. Then $\frac{h'}{q} \equiv -ah' \pmod{k}$. Hence $\frac{h'}{q}$ in (4.2) may be replaced by $-ah'$. The sum is then a Kloosterman sum with estimate $O(k^{(2/3)+\epsilon} m^{1/3})$.

We have then

$$R_1 = O\left\{\sum_{k=1}^{N} \frac{2}{kN} \sum_{n=0}^{\infty} b_n \exp\left(-\frac{\pi n}{q} + 2\pi m N^{-2}\right) k^{(2/3)+\epsilon} m^{1/3}\right\}$$
$$= O\{m^{1/3} N^{(-1/3)+\epsilon} \exp(2\pi m N^{-2})\}.$$

To find estimates for R_2 and R_3 we proceed as for S_2 and S_3. The sums in R_2 and R_3 corresponding to (4.2) will be incomplete Kloosterman sums because of the restriction placed on h' by $N < k + k_1 \leqq l$ or $N < k + k_2 \leqq l$. We find

$$\left.\begin{matrix} R_2 \\ R_3 \end{matrix}\right\} = O\{m^{1/3}N^{(-1/3)+\epsilon} \exp (2\pi m N^{-2})\}.$$

Therefore,

$$a_m^{(p)} = O\{m^{1/3}N^{(-1/3)+\epsilon} \exp (2\pi m N^{-2})\}. \tag{4.3}$$

Because of the similarity between $a_m^{(p)}$ and $a_m^{(q)}$ we know that

$$a_m^{(q)} = O\{m^{1/3}N^{(-1/3)+\epsilon} \exp (2\pi m N^{-2})\}. \tag{4.4}$$

5. Estimate for $a_m^{(1)}$. From (1.6), (2.4), (2.8) and $F(x) = \sum_{n=0}^{\infty} a_n x^n$ we get

$$\begin{aligned} a_m^{(1)} = &\sum_{\substack{k=1 \\ (k,pq)=1}}^{N} \sum_{n=0}^{\mu-1} a_n \sum_{h \bmod k}' \exp \left[-\frac{2\pi i}{k}\left(mh - \frac{nh'}{pq}\right)\right] \Omega_{h,k} \\ &\times \int_{-\theta'_{h,k}}^{\theta''_{h,k}} \exp \left[-\frac{2\pi}{k^2 pqw}(n-\mu) + 2\pi w(m-\mu)\right] d\phi \\ &+ \sum_{\substack{k=1 \\ (k,pq)=1}}^{N} \sum_{n=\mu}^{\infty} a_n \sum_{h \bmod k}' \exp \left[-\frac{2\pi i}{k}\left(mh - \frac{nh'}{pq}\right)\right] \Omega_{h,k} \\ &\times \int_{-\theta'_{h,k}}^{\theta''_{h,k}} \exp \left[-\frac{2\pi}{k^2 pqw}(n-\mu) + 2\pi w(m-\mu)\right] d\phi = I'_1 + I'_2. \end{aligned}$$

This resembles $a_m^{(pq)}$ in that the coefficient of $\dfrac{1}{w}$ in the exponent of the integrand is negative or zero for $n \geqq \mu$ and positive for $0 \leqq n \leqq \mu - 1$. Hence our treatment of $a_m^{(1)}$ is like that of $a_m^{(pq)}$. Note that h' is now a solution of $hh' \equiv -1 \pmod{k}$ chosen to be divisible by pq, and that $(k, pq) = 1$.

Corresponding to (3.3) we obtain

$$\begin{aligned} &\sum_{h \bmod k}' \exp \left[-\frac{2\pi i}{k}\left(mh - \frac{nh'}{pq}\right)\right] \Omega_{h,k} \\ &= \sum_{h \bmod k}' \exp\{-\frac{2\pi i}{k}[h(m-\mu) + \frac{h'}{pq}(-n+\mu)]\}. \end{aligned} \tag{5.1}$$

If we determine an integer b from $bpq \equiv -1 \pmod{k}$, then $\dfrac{h'}{pq} \equiv -bh'$ $\pmod{k}$. Replacing $\dfrac{h'}{pq}$ in (5.1) by $-bh'$, we see that this is a Kloosterman sum with estimate $O(k^{(2/3)+\epsilon}m^{1/3})$. Following the method used for $a_m^{(pq)}$, we obtain

$$I'_2 = O\{m^{1/3}N^{(-1/3)+\epsilon} \exp(2\pi m N^{-2})\}$$

and

$$I'_1 = 2\pi \sum_{\substack{k=1 \\ (k,pq)=1}}^{N} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m)\, L_{k,n}(m) + O\{m^{1/3}N^{(-1/3)+\epsilon} \exp(2\pi m N^{-2})\}$$

where now $B_{k,n}(m)$ is given by (5.1), $\Omega_{h,k}$ by (2.8), and

$$L_{k,n}(m) = \frac{1}{k}\sqrt{\frac{\mu - n}{pq(m-\mu)}}\, I_1\left(\frac{4\pi}{k\sqrt{pq}}\sqrt{(\mu - n)(m-\mu)}\right). \tag{5.2}$$

Again $I_1(z)$ is the Bessel function of first order with pure imaginary argument. Hence,

$$a_m^{(1)} = 2\pi \sum_{\substack{k=1 \\ (k,pq)=1}}^{N} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m) L_{k,n}(m) + O\{m^{1/3}N^{(-1/3)+\epsilon} \exp(2\pi m N^{-2})\}, \tag{5.3}$$

where $B_{k,n}(m)$ is given by (5.1), $L_{k,n}(m)$ by (5.2).

6. A formula for a_m. Collecting our results from (3.13), (4.3), (4.4), and (5.3) we may write

$$a_m = 2\pi \sum_{\substack{k=1 \\ (k,pq)=pq \\ (k,pq)=1}}^{N} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m) L_{k,n}(m) + O\{m^{1/3}N^{(-1/3)+\epsilon} \exp(2\pi m N^{-2})\}.$$

If we now let $N \to \infty$, the error term approaches 0 as a limit. Hence we obtain our formula for the number of partitions of an integer $m > \mu = \dfrac{(p-1)(q-1)}{24}$ into parts none of which are divisible by p or q:

$$a_m = 2\pi \sum_{\substack{k=1 \\ (k,pq)=pq \\ (k,pq)=1}}^{\infty} \sum_{n=0}^{\mu-1} a_n B_{k,n}(m) L_{k,n}(m),$$

where $B_{k,n}(m)$ and $L_{k,n}(m)$ are given by (3.8) and (3.9) if $(k,pq) = pq$, and by (5.1) and (5.2) if $(k,pq) = 1$. The symbols, $a_1, a_2, \cdots, a_{\mu-1}$, represent the number of partitions of $1, 2, \cdots, \mu - 1$, respectively, into summands of the type we are considering and $a_0 = 1$.

QUEENS COLLEGE,
FLUSHING, N. Y.

PARTIALLY ORDERED SETS.* [1]

By Ben Dushnik and E. W. Miller.

1. Introduction.

1. 1. By a *system* is meant a set S together with a binary relation $R(x,y)$ which may hold for certain pairs of elements x and y of S. The relation $R(x,y)$ is read "x precedes y" and is written "$x < y$." A system is called a *partial order* if the following conditions are satisfied. (1) If $x < y$, then $y \nless x$; and (2) if $x < y$ and $y < z$, then $x < z$.

A partial order defined on a set S is called a *linear order* if every two distinct elements x and y of S are comparable, i. e., if $x < y$ or $y < x$. If the partial order P and the linear order L are both defined on the same set of elements, and if every ordered pair in P occurs in L, then L will be called a *linear extension* of P.

1. 2. If P and Q are two systems on the same set of elements S, then $A = P + Q = Q + P$ will denote the system which contains those and only those ordered pairs which occur in either P or Q. Likewise $P - Q$ will denote the system which contains those and only those ordered pairs which occur in P but not in Q. The system which consists of all ordered pairs which occur in both P and Q will be denoted by $P \cdot Q$. More generally, if $P_1, P_2, \cdots, P_\alpha, \cdots$ are systems on S, then ΠP_α will denote the system which consists of all ordered pairs common to all the systems P_α. It is easily seen that ΠP_α is a partial order if each system P_α is a partial order. On the other hand, it is clear that both P and Q can be partial orders without the same being true of either $P + Q$ or $P - Q$.

2. The dimension of a partial order.

2. 1. Let S be any set, and let $\mathcal{K}$ be any collection of linear orders, each defined on all of S. We define a partial order P on S as follows. For any two elements x_1 and x_2 of S we put $x_1 < x_2$ (in P) if and only if $x_1 < x_2$ in every linear order of the collection $\mathcal{K}$; in other words, if $\mathcal{K} = \{L_\alpha\}$, we have $P = \Pi L_\alpha$. A partial order so obtained will be said to be *realized* by the linear orders of $\mathcal{K}$.

---

* Received December 8, 1940.

[1] Portions of this paper were presented to the American Mathematical Society on April 12 and November 22, 1940, under the titles "On partially ordered sets" and "On the dimension of a partial order." We are indebted to S. Eilenberg for suggestions which enabled us to simplify several proofs and improve the form of several definitions.

2. 2. By the *dimension*[2] of a partial order P defined on a set S is meant the smallest cardinal number $\mathfrak{m}$ such that P is realized by $\mathfrak{m}$ linear orders on S.

2. 3. We shall make use of the following lemma in showing that every partial order has a dimension.

LEMMA 2. 31.[3] *Every partial order P possesses a linear extension L. Moreover, if a and b are any two non-comparable elements of P, there exists an extension L_1 in which $a < b$ and an extension L_2 in which $b < a$.*

We now prove the following theorem.

THEOREM 2. 32. *If P is any partial order on a set S, then there exists a collection $\mathcal{K}$ of linear orders on S which realize P.*

Proof. If every two elements of P are comparable, then P is a linear order L and is realized by the single linear order L. If P contains non-comparable elements, then, for every non-comparable pair a and b, let $\mathcal{K}$ contain the corresponding linear extensions L_1 and L_2 mentioned in Lemma 2. 31. It is clear that P is realized by the linear orders in $\mathcal{K}$.

In light of the proof of Theorem 2. 32, the following theorem is now obvious.

THEOREM 2. 33. *Let P be any partial order on a set S. If S is finite, then the dimension of P is finite. If $\bar{\bar{S}} = \mathfrak{m}$, where $\mathfrak{m}$ is a transfinite cardinal, then the dimension of P is $\leqq \mathfrak{m}$.*

2. 4. The procedure employed in **2. 1** for defining a partial order may be formulated in the following slightly different way. Let S be any set, and let $L_1, L_2, \cdots, L_\alpha, \cdots, (\alpha < \beta)$ be a series of linear orders. (We do not require that the elements of L_α be elements of S). Let $f_1, f_2, \cdots, f_\alpha, \cdots, (\alpha < \beta)$, be a series of single-valued functions, each defined on S, each having a single-valued inverse, and such that $f_\alpha(S) \subset L_\alpha$ for every $\alpha < \beta$. We define a partial order P on S as follows. For any two elements x_1 and x_2 of S we put $x_1 < x_2$ (in P) if and only if $f_\alpha(x_1) < f_\alpha(x_2)$ for every $\alpha < \beta$. A partial order so obtained may be said to be realized by the functions f_α, and the dimension of a given partial order P may be defined as the smallest cardinal number $\mathfrak{m}$ such that P is realized by $\mathfrak{m}$ functions. It is clear that the above is nothing more than a reformulation of what appears in **2. 1** and **2. 2.**

---

[2] It will be noticed that the term "dimension" is here used in a different sense from that employed by Garrett Birkhoff in his book, *Lattice Theory*, American Mathematical Society Colloquium Publications.

[3] For a proof of this important result, see Edward Szpilrajn, "Sur l'extension de l'ordre partiel," *Fundamenta Mathematicae*, vol. 16 (1930), pp. 386-389.

3. Reversible partial orders.

3.1. Let P and Q be two partial orders on the same set of elements S, and suppose that every pair of distinct elements of S is ordered in just one of these partial orders; in such a case we shall say that P and Q are *conjugate* partial orders. A partial order will be called *reversible* if and only if it has a conjugate. Examples of reversible and irreversible partial orders will be given later.

3.2. A familiar example of a partial order is furnished by any family $\mathfrak{F}$ of subsets S of a given set E, where we put $S < S'$ if and only if S is a proper subset of S'. Conversely, if P is any partial order, then P is similar to a partial order P_1 defined as above by means of some family of sets. To show this, let us define, for any a in P, the set $S(a)$ as consisting of a and all x in P such that $x < a$. It is easily verified that $\mathfrak{F} = \{S(a)\}$, for all a in P, is the required family of sets. Any family $\mathfrak{F}$ of sets which defines (in the sense of set inclusion) a partial order similar to a given partial order P will be called a *representation* of P.

3.3. A linear extension L of P will be called *separating* if and only if there exist three elements a, b and c in P such that $a < c$, b is not comparable with either a or c in P, while in L we have $a < b < c$.

3.4. If P is a partial order, then the partial order obtained from P by inverting the sense of all ordered pairs will be denoted by P^*.

3.5. We shall need the following lemma in proving the main theorem of this section.

LEMMA 3.51. *If the partial order P has a conjugate partial order Q, then $A_1 = P + Q$ and $A_2 = P + Q^*$ are both linear extensions of P.*

The proof of this is easy, and will therefore be omitted.

3.6. We shall now prove the following theorem.

THEOREM 3.61. *The following four properties of a partial order P are equivalent.*

(1) *P is reversible.*

(2) *There exists a linear extension of P which is non-separating.*

(3) *The dimension of P is $\leqq 2$.*

(4) *There exists a representation of P by means of a family of intervals on some linearly ordered set.*

Proof. We shall show first that (1) implies (2). Suppose that the partial order P defined on the set S is reversible, and let Q be a partial order on S conjugate to P. By Lemma 3. 51, $A = P + Q$ is a linear extension of P. Let a, b and c be any three elements of S which appear in the order $a < b < c$ in A. If b is not comparable in P with either a or c, then $a < b$ and $b < c$ both appear in Q. Since Q is a partial order, $a < c$ must also appear in Q, and thus a and c are not comparable in P. Hence A is a non-separating linear extension of P.

We show now that (2) implies (3). Suppose that A is a non-separating linear extension of P, and let $Q = A - P$. If $a < b$ and $b < c$ are both in Q, then b is not comparable with either a or c in P. Then $a < c$, which is in A, cannot appear in P, for otherwise $a < b < c$ would be an instance of a separation in A. Hence $a < c$ must also appear in Q, and therefore Q is a partial order. Since Q is conjugate to P, it follows from Lemma 3. 51 that $B = P + Q^*$ is a linear extension of P, and it is obvious that P is realized by the linear orders A and B. Therefore the dimension of P is $\leqq 2$.

We prove next that (3) implies (4). Suppose the dimension of P is $\leqq 2$, and let A and B be any two linear orders on S which together realize P. (If the dimension of P is $= 1$, then P is a linear order L and we may put $A = B = L$). Let B' be a linear order similar to B^*, where the set of elements in B' is disjoint from S. Put $C = B' + A$, so that C is the linear order comprising B' and A, with the additional stipulation that each element of B' precedes each element of A. For each x in P denote by $\bar{x}$ the image of x in the given similarity transformation of B^* into B', and denote by I_x the closed interval $[\bar{x}, x]$ of C. We will show that the family $\{I_x\}$ of all such intervals is a representation of P. Suppose first that $x < y$ in P. Then $x < y$ in A and $\bar{y} < \bar{x}$ in B', so that in C we have $\bar{y} < \bar{x} < x < y$. But this means that I_x is a proper subset of I_y. In the same way it can be shown that if x and y are non-comparable elements of P, then neither of the intervals I_x and I_y contains the other.

To show that (4) implies (1) we shall suppose that P is a partial order (on a set S) which is represented by a family $\{I\}$ of intervals taken from some linear order L. For each x in S, denote by I_x the interval of the family $\{I\}$ which corresponds to x. We notice first that if x and y are distinct elements of S which are not comparable in P, then I_x and I_y cannot have the same left-hand end-point. We define a system Q on S as follows. We put $x < y$ (in Q) if and only if (a) x and y are not comparable in P, and (b) the left-hand end-point of I_x precedes the left-hand end-point of I_y. It is easy to see that Q is a partial order and that Q is conjugate to P. Hence, P is reversible.

3.7. We conclude this section with the following theorem on the question of representation.

THEOREM 3.71. *Let P be a partial order such that if $a < b$ and $a < c$, then b and c are comparable. Then P has a representation in which any two sets are either disjoint or comparable.*

Proof. Let $\mathfrak{F} \equiv \{S(x)\}$ be the representation of P defined in **3.2**. Suppose that x and y are comparable,—say $x < y$. Then clearly $S(x)$ is a proper subset of $S(y)$. If x and y are not comparable, then $S(x) \cdot S(y) = 0$. For suppose the contrary, and let $z \in S(x) \cdot S(y)$. Then $z < x$ and $z < y$. Therefore x and y are comparable, contrary to our supposition.

4. The existence of a partial order having a given dimension.

We first prove the following theorem.

THEOREM 4.1. *For every cardinal number $\mathfrak{n}$ (finite or transfinite), there exists a partial order whose dimension is $\mathfrak{n}$.*

Proof. Let X be any set of elements such that $\bar{\bar{X}} = \mathfrak{n}$. For any x in X, denote by a_x the subset of X whose only element is x, and by c_x the complement of a_x in X. Let $\mathfrak{F}$ denote the family of all sets a_x and c_x, for all x in X, and let P be the partial order *represented* (see **3.2**) by $\mathfrak{F}$. It is clear that, for any two elements a and b of P, we have $a < b$ if and only if there exist x and y in X such that $x \neq y$, $a = a_x$ and $b = c_y$. We shall prove that the dimension of P is $= \mathfrak{n}$ by showing that (1) if $x \neq y$, then no single linear extension of P can contain both $c_x < a_x$ and $c_y < a_y$; and (2) there exist $\mathfrak{n}$ linear extensions of P which realize P.

As to (1), suppose the contrary, and let L be a linear extension of P in which we have both $c_x < a_x$ and $c_y < a_y$. Since $a_x < c_y$ and $a_y < c_x$ in P, we obtain in L: $c_x < a_x < c_y < a_y < c_x$, or $c_x < c_x$, which is impossible.

As to (2), we define L_x, for any x in X, to be any specific linear extension of P in which $c_x < a_x$, and in which $a_y < c_x$ and $a_x < c_y$, for all $y \neq x$. Let $\mathfrak{K} \equiv \{L_x\}$, for all x in X. We have $\bar{\bar{\mathfrak{K}}} = \mathfrak{n}$. Moreover, the set of linear extensions $\mathfrak{K}$ realizes P. Thus, for the non-comparable elements c_x and c_y, $x \neq y$, we have $c_x < a_x < c_y$ in L_x, and $c_y < a_y < c_x$ in L_y, and similarly for the pair a_x and a_y. For the non-comparable elements a_x and c_x we have $c_x < a_x$ in L_x, and $a_x < c_y < a_y < c_x$ in any extension L_y, $y \neq x$. Finally for the comparable elements a_x and c_y, we clearly have $a_x < c_y$ in every extension L_q of $\mathfrak{K}$.

4.2. We proceed to give another example of a partial order of dimen-

sion n, for the case in which n is a finite cardinal. We shall need the following lemma.

LEMMA 4.21.[4] *For any permutation θ_n of n distinct natural numbers, there exist k of these numbers which appear in θ_n either in increasing or in decreasing order, where k is the unique natural number such that $(k-1)^2 < n \leqq k^2$.*

Now let n be any natural number and let $p = 2^{2^n} + 1$. Let S be the set whose elements are the first p natural numbers and all pairs (i, j), $i < j$, of these numbers. We define a partial order P on S as follows. If x and y are any two elements of S, let $x < y$ if and only if y is of the form (i, j) and x is either i or j. We now prove the following theorem.

THEOREM 4.22. *The partial order P just defined is of dimension $> n$.*

Proof. Let us assume that the dimension of P is $\leqq n$. There will exist n linear extensions $E_1, E_2, \cdots, E_n$ of P which realize P. The first p natural numbers, as elements of S, appear in a certain permutation in each of these linear extensions. By Lemma 4.21, we can select $2^{2^{n-1}} + 1$ of these numbers which appear monotonically (that is, in numerically increasing or decreasing order) in E_1; from these numbers we can select $2^{2^{n-2}} + 1$ which appear monotonically in E_2, etc.; so that we finally obtain $2^{2^0} + 1 = 3$ numbers which appear monotonically in every one of these linear extensions. Without loss of generality, we may suppose these numbers to be 1, 2 and 3, and that

$$(1) \qquad 1 < 2 < 3 \text{ in } E_i, \qquad (i = 1, 2, \cdots, s);$$

and

$$(2) \qquad 3 < 2 < 1 \text{ in } E_i, \qquad (i = s+1, s+2, \cdots, n).$$

Consider now the element $(1, 3)$, which follows both 1 and 3 in P. In each of the first s extensions we will have $2 < 3 < (1, 3)$, and in each of the remaining extensions we will have $2 < 1 < (1, 3)$. Hence in all of the extensions we will have $2 < (1, 3)$, so that $2 < (1, 3)$ in P. But this contradicts the definition of P. It follows, by Theorem 2.33, that there is an integer $q > n$ such that the dimension of P is $= q$.

We can now use P to obtain a partial order whose dimension is n. For let $L_1, L_2, \cdots, L_q$ be linear extensions of P which realize P. It is not hard to show that the partial order P_1 which is realized by $L_1, L_2, \cdots, L_n$, is of dimension n.

---

[4] This result appears (in slightly different form) in a paper by P. Erdös and G. Szekeres, entitled "A combinatorial problem in geometry," *Compositio Mathematica*, vol. 2 (1935), pp. 463-470.

5. Linear orders in a partial order.

5. 1. By a *graph* is meant a set of elements G, together with a binary, symmetric relation $R(x, y)$ which may hold for certain elements x and y of G. If $R(x, y)$ holds, we shall say that x and y are *connected.* A graph is said to be *complete,* if $R(x, y)$ holds for every pair of distinct elements x and y of the graph. It is clear that any partial order gives rise to a graph if we use for $R(x, y)$ the relation "x and y are comparable."

5. 2. The theorems on partial orders in this section will be obtained as consequences of certain theorems about graphs. We will prove first the following lemma.

LEMMA 5. 21. *If G is a graph of power $\mathfrak{m}$, where $\mathfrak{m}$ is a regular*[5] *cardinal, and if every subset of G of power $\mathfrak{m}$ contains two connected elements, then there exists an element x of G which is connected with $\mathfrak{m}$ elements of G.*

Proof. Assume there is no such element x. Let x_1 be any element of G, and denote by G_1 the set of all elements with which x_1 is connected. We have $\overline{\overline{G}}_1 < \mathfrak{m}$. Let μ denote the initial ordinal such that $\bar{\mu} = \mathfrak{m}$, and suppose that x_α and G_α have been defined and $\overline{\overline{G}}_\alpha < \mathfrak{m}$, for all $\alpha < \beta$, where $\beta < \mu$. Since $\mathfrak{m}$ is regular, $G - \sum_{\alpha<\beta} (x_\alpha + G_\alpha) \neq 0$. Let x_β be any element of $G - \sum_{\alpha<\beta} (x_\alpha + G_\alpha)$. Denote by G_β the set of all elements in G with which x_β is connected. Then $\overline{\overline{G}}_\beta < \mathfrak{m}$. Consider finally the set $X = \sum_{\alpha<\mu} x_\alpha$. No two elements of X are connected, and yet $\overline{\overline{X}} = \mathfrak{m}$. From this contradiction the result follows.

THEOREM 5. 22.[6] *If G is a graph of power $\mathfrak{m}$, where $\mathfrak{m}$ is a transfinite cardinal, and if every subset of G of power $\mathfrak{m}$ contains two connected elements, then G contains a complete graph of power $\aleph_0$.*

Proof. We consider first the case where $\mathfrak{m}$ is regular. In virtue of Lemma 5. 21, there exists an element x_1 of G which is connected with $\mathfrak{m}$ elements of G. Denote the set of these elements by G_1. We have $\overline{\overline{G}}_1 = \mathfrak{m}$.

---

[5] For the meaning of the terms *regular* and *singular*, in connection with transfinite numbers, one may refer to Sierpiński's book, *Leçons sur les Nombres Transfinis.* A simple type of example shows that Lemma 5. 21 is not true in case $\mathfrak{m}$ is any *singular* cardinal.

[6] We are indebted to P. Erdös for suggestions in connection with Theorems 5. 22 and 5. 23. In particular, Erdös suggested the proof of 5. 22 for the case in which $\mathfrak{m}$ is a singular cardinal.

Suppose now that x_{n-1} and G_{n-1} have been defined and $\bar{\bar{G}}_{n-1} = \mathfrak{m}$. In virtue of Lemma 5.21, there exists an element x_n of G_{n-1} such that (1) $x_n \neq x_k$, for $k < n$, and (2) x_n is connected with $\mathfrak{m}$ elements of G_{n-1}. Denote this set of elements by G_n. Consider finally the set $X = \sum_{n=1}^{\infty} x_n$. It is clear that $\bar{\bar{X}} = \aleph_0$ and that any two elements of X are connected.

We now consider the case where $\mathfrak{m}$ is a singular cardinal. Let $\mathfrak{b}$ denote the smallest cardinal such that $\mathfrak{m}$ is the sum of $\mathfrak{b}$ cardinals each less than $\mathfrak{m}$. Since $\mathfrak{m}$ is singular, we have $\mathfrak{b} < \mathfrak{m}$. Let ϕ denote the initial ordinal such that $\bar{\phi} = \mathfrak{b}$. There will exist regular cardinals $\mathfrak{r}_1, \mathfrak{r}_2, \cdots, \mathfrak{r}_\alpha, \cdots, \alpha < \phi$, such that $\mathfrak{b} < \mathfrak{r}_\alpha < \mathfrak{m}$ and $\mathfrak{m} = \sum_{\alpha < \phi} \mathfrak{r}_\alpha$.

In the first place, if every subset H of G of power $\mathfrak{m}$ contains an element connected with $\mathfrak{m}$ elements of H, then we can proceed, as in the previous case, to obtain a complete graph of power $\aleph_0$. We shall accordingly assume that there exists a subset H of G such that $\bar{\bar{H}} = \mathfrak{m}$ and such that no element of H is connected with $\mathfrak{m}$ elements of H.

We shall show that there exists an $\alpha < \phi$ and a subset Q of H such that $\bar{\bar{Q}} = \mathfrak{r}_\alpha$ and such that every subset of Q of power $\mathfrak{r}_\alpha$ contains two connected elements. Then, by Case 1, there exists in Q, and therefore in G, a complete graph of power $\aleph_0$.

Let us assume the contrary, namely, that there exists no such subset Q of H corresponding to any $\alpha < \phi$. We shall show that this assumption leads to a contradiction.

First, if A is any subset of H, denote by $C(A)$ the set of all elements of H which are connected with the various elements of A. Let K be any subset of H such that $\bar{\bar{K}} = \mathfrak{m}$. Let α be any ordinal $< \phi$. We shall show that K contains a subset W, of power $\mathfrak{r}_\alpha$, with these two properties: (1) no two elements of W are connected, and (2) $\overline{\overline{C(W)}} < \mathfrak{m}$. To prove this, we notice first that by the assumption made in the previous paragraph, there is a subset L of K such that $\bar{\bar{L}} = \mathfrak{r}_\alpha$ and such that no two elements of L are connected. Let L_β denote the set of all x in L such that x is connected with at most $\mathfrak{r}_\beta$ elements of H. We have $L = \sum_{\beta < \phi} L_\beta$. It follows that $\mathfrak{r}_\alpha = \sum_{\beta < \phi} \bar{\bar{L}}_\beta$. Since $\mathfrak{r}_\alpha$ is regular, and $\bar{\phi} = \mathfrak{b} < \mathfrak{r}_\alpha$, we must have $\bar{\bar{L}}_\beta = \mathfrak{r}_\alpha$ for some $\beta < \phi$. We now take W as this set L_β. Clearly (1) no two elements of W are connected, since $W \subset L$, and (2) $\overline{\overline{C(W)}} = \overline{\overline{C(L_\beta)}} \leqq \mathfrak{r}_\alpha \cdot \mathfrak{r}_\beta < \mathfrak{m}$.

To obtain the contradiction we proceed as follows. Denote by W_1 a subset of H such that $\bar{\bar{W}}_1 = \mathfrak{r}_1$, no two elements of W_1 are connected, and $\overline{\overline{C(W_1)}} < \mathfrak{m}$. Suppose we have defined W_α for every $\alpha < \lambda < \phi$, so that $\bar{\bar{W}}_\alpha = \mathfrak{r}_\alpha$, no two

elements of W_a are connected, and $\overline{\overline{C(W_a)}} < \mathfrak{m}$. Then $H - \sum_{a<\lambda} \{W_a + C(W_a)\}$ has power $\mathfrak{m}$. (This is the case since ϕ is the *initial* ordinal such that $\bar{\phi} = \mathfrak{b}$). Let W_λ be a subset of $H - \sum_{a<\lambda} \{W_a + C(W_a)\}$ such that $\overline{\overline{W}}_\lambda = \mathfrak{r}_\lambda$, no two elements of W_λ are connected and $\overline{\overline{C(W_\lambda)}} < \mathfrak{m}$. Now consider $\sum_{\lambda<\phi} W_\lambda$. Clearly, this set has power $\mathfrak{m}$ and yet no two elements of it are connected. But this contradicts our hypothesis.

On the basis of the theorem just proved we can now prove the following related theorem.

THEOREM 5.23. *If G is a graph of power $\mathfrak{m}$, where $\mathfrak{m}$ is a transfinite cardinal, and every subset of G of power $\aleph_0$ contains two connected elements, then G contains a complete graph of power $\mathfrak{m}$.*

Proof. Let $R'(x, y)$ mean "x and y are not connected." Let G' denote the graph determined by the elements of the set G in connection with the relation $R'(x, y)$. The application of Theorem 5.22 to the graph G' leads easily to the desired conclusion.

As previously mentioned, a partial order P gives rise to a graph if we let $R(x, y)$ mean "x and y are comparable in P." Hence the two theorems just proved give us the following theorems as corollaries.

THEOREM 5.24. *If P is a partial order of power $\mathfrak{m}$, where $\mathfrak{m}$ is a transfinite cardinal, and if every subset of P of power $\mathfrak{m}$ contains two comparable elements, then P contains a linear order of power $\aleph_0$.*

THEOREM 5.25. *If P is a partial order of power $\mathfrak{m}$, where $\mathfrak{m}$ is a transfinite cardinal, and if every subset of P of power $\aleph_0$ contains two comparable elements, then P contains a linear order of power $\mathfrak{m}$.*

5.3. The question arises as to whether stronger conclusions can be drawn in Theorems 5.22, 5.23, 5.24 and 5.25. We shall consider only a very special case of this problem; namely, the case in which $\mathfrak{m} = \aleph_1$.

Consider first the following example. Let N be any set of power $\aleph_1$. Let C denote the linear continuum, and W the well-ordered series consisting of all the ordinals of the first and second class. Let f and g denote functions (single-valued and having a single-valued inverse) defined on N and such that $f(N) \subset C$ and $g(N) \subset W$, respectively. We denote by P the (reversible) partial order on N which is realized by the two functions f and g. Now if M is any non-denumerable subset of N, there exists an element x of M such that $f(x)$ is a condensation point of $f(M)$ from both the left and the right. There accordingly exist elements y and z of M such that $f(z) < f(x) < f(y)$, $g(x) < g(y)$, and $g(x) < g(z)$. Hence, x and y are comparable in P, while

x and z are not comparable in P.[7] In other words, the partial order P has the following property: Every subset of P of power $\aleph_1$ contains two comparable elements, and yet P contains no linear order of power $\aleph_1$.

Since P is reversible, it can (by virtue of Theorem 3. 61) be represented as a family of intervals on some linear order. Hence, the result just obtained can be given the following form.

THEOREM 5. 31. *There exists a non-denumerable family $\mathfrak{F}$ of intervals (on a certain linear order A of power $\aleph_1$) which has the following property. Every non-denumerable sub-family $\mathfrak{F}'$ of $\mathfrak{F}$ contains two comparable intervals, and yet $\mathfrak{F}$ contains no non-denumerable monotonic sub-family.*

A stronger result than that of Theorem 5. 31 will be presently obtained. This result will depend upon the following theorem.

THEOREM 5. 32. *If the hypothesis of the continuum is true, there exists a non-denumerable set N of real numbers which has the following property. If N_1 and N_2 are any two disjoint non-denumerable subsets of N, then $\phi(N_1) \neq N_2$, where ϕ is any increasing or decreasing function defined on N_1.*

Proof. Let us arrange in a well-ordered series of type Ω all real-valued functions $f(x)$ which (a) are monotonic (non-increasing and non-decreasing) on the linear continuum C, and (b) are such that $E[f(x) = x]$ is nowhere dense on C:

$$f_1, f_2, \cdots, f_\alpha, \cdots, (\alpha < \Omega).$$

For a given $\alpha < \Omega$ there may exist an interval on which $f_\alpha(x)$ is constant. The set of all values assumed on such intervals (for a given α) is at most denumerably infinite. Denote this set of values by D_α.

Let x_1 be any real number, and assume that x_β has been defined for all $\beta < \alpha < \Omega$. We shall show that it is possible to choose x_α so that (1) $x_\alpha \neq x_\beta$ for $\beta < \alpha$, (2) $x_\alpha \neq f_\mu(x_\beta)$ for $\mu < \alpha$ and $\beta < \alpha$, (3) $f_\mu(x_\alpha) \neq x_\alpha$ for $\mu < \alpha$, and (4) $f_\mu(x_\alpha) \neq x_\beta$ or $f_\mu(x_\alpha) \in D_\mu$ for $\mu < \alpha$ and $\beta < \alpha$. That x_α can be so defined may be seen as follows. Conditions (1) and (2) can be realized by avoiding a denumerable set. By virtue of (b), we can realize (3) by avoiding a set of the first category. Finally consider any $\mu < \alpha$ and any $\beta < \alpha$. There is at most one x in $C - D_\mu$ for which $f_\mu(x) = x_\beta$. It follows that, except for a countable set of points, we have $f_\mu(x) \in D_\mu$ or $f_\mu(x) \neq x_\beta$ for all $\mu < \alpha$ and all $\beta < \alpha$. Altogether, then, the set of points which has to

[7] In a similar way it can be shown that if N_1 and N_2 are any two disjoint non-denumerable subsets of N, then there exist elements a_1 and b_1 of N_1, and elements a_2 and b_2 of N_2, such that a_1 and a_2 are comparable in P, while b_1 and b_2 are not comparable in P.

be avoided is of the first category. As such a set cannot exhaust C, it is clear that (1), (2), (3) and (4) can be realized.

We now put $N = \sum_{\alpha < \Omega} x_\alpha$. From (1) it follows that N is non-denumerable. Consider now any fixed $\mu < \Omega$, and any λ such that $\mu < \lambda < \Omega$. From (2), (3) and (4) it can be seen that if $f_\mu(x_\lambda) \in N$, then $f_\mu(x_\lambda) \in D_\mu$. It follows that $N \cdot f_\mu(N)$ is denumerable. Hence, for no $\mu < \Omega$ can we have $f_\mu(N_1) = N_2$, where N_1 and N_2 are disjoint non-denumerable subsets of N. Finally, assume that $\phi(N_1) = N_2$, where ϕ is an increasing or decreasing function defined on N_1. For each x in N_1, we have $\phi(x) \neq x$, and it can be easily shown (by suitably extending the definition of ϕ) that there exists a $\mu < \Omega$ such that f_μ agrees with ϕ on N_1. Our result follows from this contradiction.

The result of the preceding theorem can be expressed by saying that if N_1 and N_2 are disjoint non-denumerable subsets of N, then N_1 cannot be mapped onto N_2 by any order-preserving or order-reversing transformation. It follows of course that if N_1 and N_2 are non-denumerable subsets of N such that $N_1 - N_2$ is non-denumerable, then N_1 cannot be mapped onto N_2, for such a mapping would imply that $N_1 - N_2$ could be mapped onto a non-denumerable subset of N_2. In a previous paper[8] the authors have shown that there exists a non-denumerable subset of the linear continuum which is not similar to any proper subset of itself. We note here that the set N just constructed has the following property: If M is any non-denumerable subset of N, then M is not similar to any proper subset of itself which differs from M in more than a denumerable infinity of points.

We now return to our main purpose, and prove the following theorem.

THEOREM 5.33. *The set N of Theorem 5.32 has the following property. Let $\mathfrak{F}$ be <u>any</u> non-denumerable family of intervals on the linearly ordered set N such that no two intervals of $\mathfrak{F}$ have an end-point in common. Then $\mathfrak{F}$ contains two comparable intervals and two non-comparable intervals.*

Proof. Assume that every two intervals of $\mathfrak{F}$ are comparable. Let N_1 denote the set of left-hand end-points and N_2 the set of right-hand end-points of the intervals of $\mathfrak{F}$. If $n_1' < n_1''$, then $n_2'' < n_2'$, and we obtain an order-reversing transformation of N_1 into N_2. Similarly, if we assume that no two intervals of $\mathfrak{F}$ are comparable, we obtain an order-preserving transformation of N_1 into N_2.

UNIVERSITY OF MICHIGAN.

[8] "Concerning similarity transformations of linearly ordered sets," *Bulletin of the American Mathematical Society*, vol. 46 (1940), pp. 322-326.

ON NON-EQUIDISTRIBUTED AVERAGES.*

By Richard Kershner.

1. The following theorem was given by F. John:[1]

Theorem. *If $f(x)$ is a periodic function of bounded variation with the period 1, and if $\gamma = p/q > 1$ is a given rational number, $(p, q) = 1$, then*

$$(1) \qquad \sum_{n=1}^{\infty} \frac{a_n(\gamma)}{n} f\left(x - \frac{\log n}{\log \gamma}\right) = \log \gamma \int_0^1 f(y)\,dy,$$

where $a_n(\gamma)$ is defined by

$$(2) \qquad a_n(\gamma) = \begin{cases} 0; & p \nmid n, q \nmid n \\ -p; & p \mid n, q \nmid n \\ q; & p \nmid n, q \mid n \\ q - p; & p \mid n, q \mid n \end{cases}$$

The problem of establishing the identity (1) for any function $f(x)$ of period 1 reduces[2] very easily to the problem of showing that

$$(3) \qquad \lim \sum_{m=r+1}^{r\gamma} \frac{1}{m}\, g\left(\frac{\log m}{\log \gamma}\right) = \log \gamma \int_0^1 g(y)\,dy,$$

where $g(y) = f(-y)$ is again periodic of period 1. It was suggested by Wintner that the limiting relation (1), whose essential content is the identity (3), has nothing to do with the ζ-function of Riemann or its generalizations as considered *loc. cit.*[1] but is only a special manifestation of some quite general principle connected with the approximation of integrals by non-equidistributed Riemann sums. It turns out that this is the case, and indeed that the general principle involved is quite trivial. However, though several special cases have been given, it would appear that the general rule has been overlooked. Thus, an interesting identity of Pólya containing the "Anzahlfunktion" of an increasing sequence is another manifestation of the general rule to be given. In § 2 this rule is given and then some of the special cases that occur in the literature are indicated in § 3 and § 4.

---

* Received December 2, 1940.

[1] F. John, "Identitäten zwischen dem integral einer willkürlichen Funktion und unendlichen Reihen," *Mathematische Annalen*, vol. 110 (1935), pp. 718-721.

[2] Cf. H. Rademacher, "Some remarks on F. John's identity," *American Journal of Mathematics*, vol. 58 (1936), p. 171.

2. Let $\{x_j^r\}$, $j = 1, 2, \cdots, n_r$, $r = 1, 2, \cdots$ (r not an exponent) be an infinite sequence of finite sequences such that

$$0 \leqq x_1^r \leqq x_2^r \leqq \cdots \leqq x_{n_r}^r \leqq 1. \tag{4}$$

The monotone non-decreasing step-function $\phi_r(x)$ defined, for $0 \leqq x \leqq 1$, by

$$\phi_r(x) = (\text{number of values of } j \text{ for which } x_j^r \leqq x)/n_r \tag{5}$$

is called the distribution function of the sequence $\{x_j^r\}$ for a fixed r. If there is a function $\phi(x)$ such that

$$\phi_r(x) \to \phi(x) \quad \text{as } r \to \infty$$

at those points x at which $\phi(x)$ is continuous, then $\phi(x)$ is called the asymptotic distribution function of $\{x_j^r\}$. Clearly $\phi(x)$ is monotone non-decreasing and

$$\int_0^1 d\phi(x) = 1. \tag{6}$$

It is a well-known consequence of this definition that

$$\lim_{r\to\infty} \frac{1}{n_r} \sum_{j=1}^{n_r} f(x_j^r) = \int_0^1 f(y)\,d\phi(y) \tag{7}$$

whenever the integral on the right exists in the Riemann-Stieltjes sense.

Now suppose that the asymptotic distribution function $\phi(x)$ is absolutely continuous and possesses a density $\delta(x)$ which does not vanish in $[0, 1]$. Then application of the relation (5) to the function $f(y) = g(y)/\delta(y)$ gives the following

THEOREM. *Suppose that $\{x_j^r\}$ satisfies* (4) *and has an absolutely continuous asymptotic distribution function with a density $\delta(x)$ which is Riemann integrable and does not vanish*[3] *in* $[0, 1]$. *Then*

$$\lim_{r\to\infty} \frac{1}{n_r} \sum_{j=1}^{n_r} g(x_j^r)/\delta(x_j^r) = \int_0^1 g(y)\,dy$$

for any function $g(y)$ which is Riemann-integrable on $[0, 1]$.

3. Consider the special case

$$x_j^r = (\log(r + j) - \log r)/\log\gamma, \qquad (j = 1, 2, \cdots, [(\gamma - 1)r]) \tag{8}$$

where $\gamma > 1$ is fixed.[4] Then (4) is clearly satisfied with $n_r = [(\gamma - 1)r]$.

---

[3] Notice that there is some latitude in the choice of $\delta(x)$. Here, any choice of $\delta(x)$ which does not violate the Riemann integrability requirement is legitimate.

[4] γ can be irrational.

Also,

$$\phi_r(x) = \text{number of values } j \text{ for which } \log(r+j) < (\log r + x \log\gamma)/[(\gamma-1)r]$$
$$= [r(e^{x\log\gamma}-1)]/[(\gamma-1)r].$$

Thus there is an asymptotic distribution

$$\phi(x) = \lim_{r\to\infty}\phi_r(x) = (e^{x\log\gamma}-1)/(\gamma-1)$$

which is absolutely continuous with the continuous density

$$\delta(x) = \log\gamma\, e^{x\log\gamma}/(\gamma-1)$$

which does not vanish in $[0,1]$. Notice that

$$\delta(x_j{}^r) = \frac{(r+j)\log\gamma}{r(\gamma-1)}.$$

Thus, according to the preceding theorem,

$$\lim_{r\to\infty}\frac{(\gamma-1)r}{[(\gamma-1)r]\log\gamma}\sum_{j=1}^{[(\gamma-1)r]}\frac{1}{r+j}\,g\left(\frac{\log(r+j)}{\log\gamma}-\frac{\log r}{\log\gamma}\right)=\int_0^1 g(y)\,dy.$$

Clearly this limit holds not only for integer values of r. In particular if r tends to ∞ in such a way that $\log r/\log\gamma$ is an integer and if $g(y)$ is defined outside $[0,1]$ to be periodic with period 1, we have

$$\lim_{r\to\infty}\frac{1}{\log\gamma}\sum_{j=1}^{[(\gamma-1)r]}\frac{1}{r+j}\,g\left(\frac{\log(r+j)}{\log\gamma}\right)=\int_0^1 g(y)\,dy.$$

Again this limit clearly holds for any sequence of r-values. This establishes the limit relation (3) and in turn the identity (1) of F. John, for any Riemann integrable function which is of period 1 on $[-\infty,\infty]$.

4. Let

$$q_1 < q_2 < \cdots < q_j < \cdots$$

be a sequence of real numbers. Then the function

$$N(r) = (\text{number of values of } j \text{ for which } q_j \leqq r)$$

is the enumerating function (Anzahlfunktion) of $\{q_j\}$. Let

$$x_j{}^r = q_j/r \qquad (j=1,2,\cdots,N(r)). \tag{9}$$

Then (4) is clearly satisfied with $n_r = N(r)$. Also

$$\phi_r(x) = (\text{number of values of } j \text{ for which } q_j \leqq rx)/N(r) \tag{10}$$
$$= N(rx)/N(r).$$

Thus there is an asymptotic distribution function for $\{x_j{}^r\}$ if and only if the monotone function $N(r)$ is such that

$$\lim N(rx)/N(r) \text{ exists for } 0 \leqq x \leqq 1.$$

According to Karamata,[5] this is the case if and only if

$$N(r) = r^{\alpha}L(r), \qquad \alpha > 0 \tag{11}$$

for some positive α, where $L(r)$ is a "slow" function, i. e., where $L(tr) \sim L(r)$, as $r \rightarrow \infty$, for every t. Suppose then that (11) holds. Then, by (10), there is an asymptotic distribution

$$\phi(x) = \lim \phi_r(x) = x^{\alpha}$$

which is absolutely continuous, with a density

$$\delta(x) = \alpha x^{\alpha-1}$$

which does not vanish on $(0, 1]$. One can define $\delta(0) = 1$ without violation of the Riemann integrability requirement and then the preceding theorem is applicable. Thus we have, by (9),

$$\lim_{r\rightarrow\infty} \frac{r^{\alpha-1}}{\alpha N(r)} \sum_{q_j \leqq r} \frac{1}{q_j{}^{\alpha-1}} f\left(\frac{q_j}{r}\right) = \int_0^1 f(x)\,dx$$

for any Riemann integrable function $f(x)$ on $[0, 1]$.

This limit is very closely related to one given by Pólya,[6] under the same assumption (11); it differs only in the fact that Pólya first performs a change of variable in the integral on the right.

THE JOHNS HOPKINS UNIVERSITY.

[5] J. Karamata, "Sur un mode de croissance régulière," *Bulletin de la Société Mathématique de France*, vol. 61 (1933), pp. 55-62.

[6] G. Pólya, "Bemerkungen über unendliche Folgen und ganze Funktionen," *Mathematische Annalen*, vol. 88 (1923), p. 173.

ON THE DECOMPOSITION OF A HILBERT SPACE BY ITS HARMONIC SUBSPACE.*

By Alexander Weinstein.

Let $\mathfrak{H}$ be the Hilbert space of all functions of integrable square defined in a domain S in 3-dimensional space. $\mathfrak{H}$ contains the subspace $\mathfrak{P}$ of all harmonic functions regular in S and of integrable square in this domain. Let $\mathfrak{Q}$ denote the subspace of $\mathfrak{H}$ of functions orthogonal to $\mathfrak{P}$. The decomposition $\mathfrak{H} = \mathfrak{P} + \mathfrak{Q}$ was investigated in a joint paper of N. Aronszajn and the author some years ago in a problem of the unified theory of eigenvalues of plates and membranes.[1] The results there obtained lead easily to a demonstration of a lemma which plays a central rôle in a recent paper of H. Weyl.[2]

Let the domain S be a sphere with the boundary C. Let K_2 be the class of all continuous functions ζ possessing continuous derivatives of the first and second order in $S + C$ and vanishing together with all these derivatives on C. Denoting by (v, w) the scalar product of two functions in $\mathfrak{H}$, we can formulate Weyl's lemma 2 in the following way.

Lemma. *A function η in $\mathfrak{H}$ satisfying the equation*

$$(\eta, \Delta\zeta) = 0 \tag{1}$$

for every ζ in K_2 is a harmonic function.

Weyl used for his proof a modification of a special construction which he ascribed to Fubini, Courant and Friedrichs, and which, together with another construction, he considered to be the backbone of his paper. The present proof is based on the use of a complete sequence in $\mathfrak{Q}$ of functions $f_1, f_2, \cdots$ which are orthogonal to η. This sequence was already introduced in our previous paper, so that we shall have to prove only such properties as were not developed there.

---

* Received March 20, 1941; Presented to the American Mathematical Society, February 22, 1941.

[1] N. Aronszajn et A. Weinstein, *Comptes rendus Acad. Sc.*, Paris, vol. 204 (1937), p. 96; N. Aronszajn and A. Weinstein, "Existence, Convergence and Equivalence in the Unified Theory of Plates and Membranes," *Proceedings of the National Academy of Sciences*, vol. 27 (1941), pp. 188-191.

[2] H. Weyl, "The Method of orthogonal projection in potential theory," *Duke Mathematical Journal*, vol. 7 (1940), pp. 411-444.

Let us denote, as in our previous paper, by Gf the Green's potential of a function f in $\mathfrak{H}$ (i. e., Gf is the formal solution, defined by means of the Green's function, of the Poisson equation $\Delta w = f$ for w vanishing on the boundary C). G is a symmetric operator in $\mathfrak{H}$. We define f_n $(n = 1, 2, \cdots)$ as a solution of the following

PROBLEM P_n. *Find a minimizing function for the expression*

$$(f, f)/(Gf, Gf)$$

for all f in $\mathfrak{Q}$. For $n > 1$, f has to satisfy the additional orthogonality conditions

$$(2) \qquad (Gf, Gf_k) = (f, GGf_k) = 0 \quad \text{for} \quad (k = 1, 2, \cdots, n-1),$$

where f_k is a solution of P_k. We denote by λ_n the minimum in P_n.

It can be proved in the usual way, that $(f_n, f_m) = 0$ for $n \neq m$ and it can be assumed that $(f_n, f_n) = 1$. In order to prove the completeness of our sequence in $\mathfrak{Q}$ we denote by ψ an arbitrary function in $\mathfrak{Q}$. Put $c_m = (\psi, f_m)$ and

$$(3) \qquad \psi_n = \psi - \sum_{i=1}^{n} c_i f_i.$$

We then have

$$(\psi_n, f_k) = 0 \quad \text{and} \quad (G\psi_n, Gf_k) = 0 \quad \text{for} \quad (k = 1, 2, \cdots, n),$$

so that

$$(4) \qquad (\psi_n, \psi_n) = (\psi, \psi) - \sum_{i=1}^{n} c_i^2$$

and

$$(5) \qquad (G\psi_n, G\psi_n) = (G\psi, G\psi) - \sum_{i=1}^{n} \frac{c_i^2}{\lambda_i}.$$

Since ψ_n is an admissible function in P_{n+1}, we have

$$(G\psi_n, G\psi_n) \leqq \frac{1}{\lambda_{n+1}} (\psi_n, \psi_n) = \frac{1}{\lambda_{n+1}} [(\psi, \psi) - \sum_{i=1}^{n} c_i^2] \leqq \frac{(\psi, \psi)}{\lambda_{n+1}}.$$

Since

$$\lambda_{n+1} \to \infty \quad \text{as} \quad n \to \infty,$$

we have

$$(6) \qquad (G\psi_n, G\psi_n) \to 0.$$

From this *strong* convergence we can conclude the *weak* convergence $(G\psi_n, \chi) \to 0$ for every χ in $\mathfrak{H}$. Using the symmetry of G, we have also $(\psi_n, G\chi) \to 0$. Since every function f in $\mathfrak{H}$ can obviously be approximated in

the sense of the metric (f, f) by functions vanishing on C and possessing a continuous Laplacian, f can be approximated by functions of the type $G\chi$. Therefore we have finally

$$(\psi_n, f) \rightarrow 0 \tag{7}$$

for every f in $\mathfrak{H}$.

From (4) we have for $n > m$

$$(\psi_n - \psi_m, \psi_n - \psi_m) = \sum_{i=m}^{n} c_i^2 \rightarrow 0 \quad \text{as} \quad n, m \rightarrow \infty.$$

Consider now

$$(\psi_n, \psi_n) = (\psi_n - \psi_m, \psi_n - \psi_m) + 2(\psi_n, \psi_m) - (\psi_m, \psi_m).$$

we have

$$(\psi_n, \psi_n) \leqq (\psi_n - \psi_m, \psi_n - \psi_m) + 2\,|(\psi_n, \psi_m)|$$

and, by using (7) applied to $f = \psi_m$, we see, in the usual way, that $(\psi_n, \psi_n) \rightarrow 0$, so that we obtain from (4) the equation $(\psi, \psi) = \sum_{i=1}^{\infty} c_i^2$ which expresses the completeness of the sequence $f_1, f_2, \cdots$ in $\mathfrak{Q}$.[3]

Let us now consider the functions w_n defined by the equations $\Delta w_n = f_n$ in S, $w_n = 0$ on C. By (2) we have $(w_n, w_m) = (Gf_n, Gf_m) = 0$ for $n \neq m$. It has been proved in our previous papers that the w_n are the eigenfunctions, possessing continuous first and second derivatives, of the equation

$$\Delta\Delta w = \lambda w \text{ in } S \tag{8}$$

which satisfy the conditions

$$w = 0 \text{ on } C, \qquad \Delta w \text{ is in } \mathfrak{Q} \tag{9}$$

and correspond to the eigenvalues $\lambda = \lambda_n$.

Let us consider, on the other hand, the eigenfunctions w^*_n of the equation (8) satisfying the classical boundary conditions

$$w = dw/d\nu = 0 \text{ on } C \tag{10}$$

where ν denotes the normal. These functions are explicitly known and, from their expression, it is evident that they form a complete orthogonal set in $\mathfrak{H}$. We now prove that the w^*_n satisfy the conditions (9). In fact, we have, by Green's formula, $(p^*, \Delta w^*_n) = 0$ for every harmonic function p^* which is regular in a domain containing $S + C$. Since every p in $\mathfrak{P}$ can be approximated in the sense of the metric in $\mathfrak{H}$ by functions of the type p^*,[4] we have

[3] This proof holds for any domain admitting a Green's function.

[4] This fact was already explicitly stated for a circle as long ago as 1907 by S. Zaremba, "L'équation biharmonique et une classe remarquable de fonctions fondamentales harmoniques," *Bull. Acad. Sc. de Cracovie* (1907), pp. 147-196 (see p. 161).

also $(p, \Delta w^*_n) = 0$, so that Δw^*_n is in $\mathfrak{Q}$. For this reason the w^*_n form in $\mathfrak{H}$ a *complete* orthogonal set of eigenfunctions of (8), (9), so that, for every eigenvalue, the w_n are linear combinations of the w^*_n.[5] The w_n and the Δw_n are therefore also of the same form as the w^*_n, namely $\Phi(r)\Psi(\alpha, \theta)$, where r, α, θ denote polar coördinates.

We can now prove Weyl's lemma in the following way. The eigenfunctions w_n are not in K_2, since $d^2w_n/d\nu^2$ does not vanish on C. However, it is, in view of (10), obvious that $\Delta w_n = f_n$ can be approximated in the sense of the metric (f, f) by functions $\Delta\zeta$ where ζ is in K_2. We have therefore

$$(\eta, f_n) = 0 \qquad (n = 1, 2, \cdots). \tag{11}$$

Splitting η into the sum $\phi + \psi$ where ϕ is in $\mathfrak{P}$ and ψ in $\mathfrak{Q}$, we have $(\eta, f_n) = (\psi, f_n) = 0$ for $n = 1, 2, \cdots$. Since the set $f_1, f_2, \cdots$ is complete in $\mathfrak{Q}$, we have $\psi = 0$ and $\eta = \phi$ so that η is a harmonic function in S.

22 West 75th Street,
New York City.

[5] Due to the fact that S is a sphere, we have here avoided the use of a general equivalence theorem proved in our previous papers.

ON THE LATTICE PROBLEM OF GAUSS.*

By Aurel Wintner.

Let $r(n)$ denote the number of solutions of the Diophantine equation $\mu^2 + \nu^2 = n$ in real integers μ, ν, and let

$$R(x) = \sum_{n \leqq x} r(n), \quad \text{where} \quad x \geqq 1;$$

so that $R(x)$ is the number of points within or on the circle of radius $x^{\frac{1}{2}}$ about the point $(0, 0)$ of a quadratic lattice of unit width. It is known that $R(x)$ not only is asymptotically proportional to the area πx, but, as shown by Gauss,

$$R(x) = \pi x + O(x^{\frac{1}{2}}); \qquad x \rightarrow \infty.$$

To-day it is known that, on the one hand,

$$R(x) = \pi x + O(x^{\alpha}) \text{ for certain } \alpha < 1/3,$$

and that, on the other hand,

$$R(x) = \pi x + \Omega_{\pm}(x^{1/4}).$$

Obviously, the jumps of the step-function $R(x)$, $1 \leqq x < \infty$, take place when x passes through an integer. If $\bar{R}(x)$ denotes the arithmetical mean of the right- and left-hand limits,

$$2\bar{R}(x) = R(x+0) + R(x-0),$$

then, according to Hardy,[1]

$$\bar{R}(x) = \pi x + x^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{r(n)}{n^{\frac{1}{2}}} J_1(2\pi(nx)^{\frac{1}{2}}),$$

for every $x > 1$, where $J_1(t)$ denotes the Bessel function of the first kind and of order 1. Furthermore, the partial sums of the convergent series on the right are uniformly bounded on every bounded interval of the half-line $x \geqq 1$. Finally, by the asymptotic formula for $J_1(t)$,

$$\bar{R}(x) = \pi x + \frac{x^{1/4}}{\pi} \sum_{n=1}^{\infty} \frac{r(n)}{n^{3/4}} \sin\left(2\pi(nx)^{\frac{1}{2}} - \tfrac{1}{4}\pi\right) + O(1), \qquad (x \rightarrow \infty),$$

where the series on the right is again boundedly convergent for $1 \leqq x <$ const.

---

* Received January 8, 1941.

[1] G. H. Hardy, "On the expression of a number as the sum of two squares," *Quarterly Journal of Mathematics*, vol. 46 (1915), pp. 263-283.

It will be convenient to denote the radius $x^{\frac{1}{2}}$ by t; so that $t \geqq 1$ and

$$(1) \qquad Q(t) = \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{r(n)}{n^{3/4}} \sin (2\pi n^{\frac{1}{2}} t - \tfrac{1}{4}\pi) + O(t^{-\frac{1}{2}}), \qquad (t \to \infty),$$

where $Q(t)$ is an abbreviation for the reduced deviation of the exact value, $\bar{R}(t^2)$, from the circular area, πt^2; so that

$$(2) \qquad Q(t) = \frac{\bar{R}(t^2) - \pi t^2}{t^{\frac{1}{2}}}.$$

In view of (1), it is natural to ask in what sense, if any, is the convergent *trigonometric* series (1) the *Fourier* series of the reduced remainder term, (2). The object of this paper is to answer this question, by proving that (2) *is almost periodic* (B^2), $1 \leqq t < \infty$, *and has precisely the Fourier series which one would expect to belong to* (2) *in virtue of* (1); *so that*

$$(3) \qquad Q(t) \sim \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{r(n)}{n^{3/4}} \sin (2\pi n^{\frac{1}{2}} t - \tfrac{1}{4}\pi) \qquad (B^2).$$

This, but not the explicit convergent representation (1), implies, in particular, the existence of an asymptotic distribution function for the reduced remainder term (2).

Incidentally, it is clear from (2) and from the connection between $\bar{R}(t^2)$ and $R(t^2)$, that (3) is equivalent to

$$\frac{\bar{R}(t^2) - \pi t^2}{t^{\frac{1}{2}}} \sim \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{r(n)}{n^{3/4}} \sin (2\pi n^{\frac{1}{2}} t - \tfrac{1}{4}\pi) \qquad (B^2).$$

If $M\{f(t)\}$ denotes the average

$$\lim_{T\to\infty} T^{-1} \int_1^T f(t)\,dt,$$

the assertion (3) may be formulated as follows: The average

$$M\{(Q(t) - \frac{1}{\pi} \sum_{n=1}^{m} \frac{r(n)}{n^{3/4}} \sin (2\pi n^{\frac{1}{2}} t - \tfrac{1}{4}\pi))^2\}$$

exists for every fixed m and tends to 0 as $m \to \infty$. Hence, it is clear from Bessel's identity that (3) will be proved if and only if one ascertains the following two facts:

(i) For every real λ, the Fourier constant $M\{e^{i\lambda t}Q(t)\}$ exists and

$$(4) \qquad M\{e^{i\lambda t}Q(t)\} = \begin{cases} 0 \text{ unless } \lambda = \pm 2\pi n^{\frac{1}{2}}, \\ \dfrac{\pm i}{2\pi} \dfrac{r(n)}{n^{3/4}} \exp(\pm \tfrac{1}{4} i\pi) \text{ for } \lambda = \pm 2\pi n^{\frac{1}{2}}, \end{cases}$$

where $n = 1, 2, \cdots$;

(ii) the square average of $Q(t)$ exists and is equal to the square sum of the amplitudes,

$$M\{Q(t)^2\} = \frac{1}{2\pi^2} \sum_{n=1}^{\infty} \frac{r(n)^2}{n^{3/2}}. \tag{5}$$

(The convergence of the series on the right of (5) is clear from the estimate

$$r(n) = O(n^\epsilon), \qquad n \to \infty, \tag{6}$$

which follows for every fixed $\epsilon > 0$ from the definition of $r(n)$).

A result much weaker than (ii) is known. In fact, Cramér [2] inferred from the results of Hardy [1] that

$$\frac{1}{X^{3/2}} \int_1^X (\bar{R}(x) - \pi x)^2 dx \to \frac{1}{3\pi^2} \sum_{n=1}^{\infty} \frac{r(n)^2}{n^{3/2}} \text{ as } X \to \infty.$$

According to (2), this can be written in the form

$$\frac{1}{T} \int_1^T Q(t)^2 \frac{t^2}{T^2} dt \to \frac{1}{6\pi^2} \sum_{n=1}^{\infty} \frac{r(n)^2}{n^{3/2}}; \qquad T \to \infty. \tag{5 bis}$$

In view of the weight factor t^2/T^2 on the left, (5 bis) is a trivial Abelian consequence of (5). In fact, two partial integrations show that

$$\text{if } \frac{1}{T}\int_0^T p(t)\,dt \to A, \text{ then } \frac{1}{T}\int_0^T p(t)\frac{t^2}{T^2}\,dt \to \frac{A}{3}; \qquad T \to \infty.$$

On the other hand, obvious examples show that the converse of this Abelian lemma is false, $p(t) \geqq 0$ being, in the present case, an insufficient restriction for Tauberian purposes.

Since it does not appear to be possible to ascertain a general Tauberian theorem leading from (5 bis) to (5), it will be necessary to leave (5 bis) aside and to prove (ii) by a suitable modification of the proof of (5 bis). On the other hand, (i) will be proved as a consequence of bounded convergence. It should be mentioned that this arrangement can be applied conveniently also to problems of almost periodicity treated earlier by direct procedures.[3]

For every $T > 1$ and for every pair of positive integers m, n, put

$$K_{mn}(T) = \int_1^T t^{-1}(f_n(t+1) - f_n(t))(f_m(t+1) - f_m(t))\,dt, \tag{7}$$

[2] H. Cramér, "Ueber zwei Sätze des Herrn G. H. Hardy," *Mathematische Zeitschrift*, vol. 15 (1922), pp. 201-210.

[3] A. Wintner, "On the asymptotic distribution of the remainder term of the prime-number theorem," *American Journal of Mathematics*, vol. 57 (1935), pp. 534-538; "The almost periodic behavior of the function $1/\zeta(1+it)$," *Duke Mathematical Journal*, vol. 2 (1936), pp. 443-446.

where

$$f_n(t) = t^{3/4} \cos (2\pi n^{\frac{1}{2}} t^{\frac{1}{2}} - \tfrac{1}{4}\pi). \tag{8}$$

It is then easy to verify, by standard applications of the second mean-value theorem, that [4]

$$| K_{mn}(T) | < \frac{\text{const.}}{m^{\frac{1}{2}} - n^{\frac{1}{2}}} \operatorname{Min} (m^{\frac{1}{4}} n^{\frac{1}{4}} \log T, T + T^{\frac{1}{2}} m^{\frac{1}{2}}) \text{ if } m > n, \tag{9}$$

where const. is independent of m, n and $T (> 2$, say$)$; that

$$K_{nn}(T) = \pi^2 n T^{\frac{1}{2}} + O(n^{3/2}) + O(n^{\frac{1}{2}} \log T) \text{ as } n \to \infty,\ T \to \infty, \tag{10}$$

where the first O-term is uniform in T; finally, that

$$K_{nn}(T) = O(T^{3/2}) \text{ as } T \to \infty, \text{ uniformly in } n. \tag{11}$$

Actually, these elementary estimates result immediately if one adapts to the present purpose the inequalities used by Cramér (*loc. cit.*[2], pp. 205-206) in the proof of (5 bis); his inequalities being adjusted to the relation

$$\bar{R}(x) - \pi x = -\frac{1}{\pi^2} \sum_{n=1}^{\infty} \frac{r(n)}{n^{5/4}} f_n(x) + O(x^{\epsilon}), \qquad x \to \infty, \tag{12}$$

which is, for every fixed $\epsilon > 0$, an immediate consequence of Hardy's explicit formula (cf. *loc. cit.*[2], p. 204).

It is clear from (6) and (8) that the series

$$S(t) = -\frac{1}{\pi^2} \sum_{n=1}^{\infty} \frac{r(n)}{n^{5/4}} f_n(t) \tag{13}$$

is absolutely-uniformly convergent on any fixed t-interval $1 \leqq t \leqq T$. Hence, if use is made of the abbreviation (7), term-by-term integration of the square of (13) gives

$$\int_1^T \frac{S(t)^2}{t} dt = \frac{1}{\pi^4} \sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{r(m) r(n)}{m^{5/4} n^{5/4}} K_{mn}(T), \tag{14}$$

where the double series on the right is absolutely convergent for every T. For every fixed T, break this double series into four parts, $L_k = L_k(T)$, as follows:

$$\begin{aligned} &\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} \frac{r(m) r(n)}{m^{5/4} n^{5/4}} K_{mn}(T) \\ &= \sum_{\substack{m=n \\ m \leqq T}} + \sum_{\substack{m=n \\ m > T}} + 2\sum_{\substack{n < m \\ m \leqq T}} + 2\sum_{\substack{n < m \\ m > T}} \\ &= L_1(T) + L_2(T) + L_3(T) + L_4(T). \end{aligned} \tag{15}$$

---

[4] Cf., e. g., N. Wiener and A. Wintner, "Fourier-Stieltjes transforms and singular infinite convolutions," *American Journal of Mathematics*, vol. 60 (1938), pp. 512-522 (more particularly, pp. 516-517).

These four functions of T will now be dealt with separately.

First, from (15) and (10),

$$L_1(T) = \sum_{n \leqq T} \frac{r(n)^2}{n^{5/2}} (\pi^2 n T^{\frac{1}{2}} + O(n^{3/2}) + O(n^{\frac{1}{2}} \log T)).$$

Hence, for every $\epsilon > 0$,

$$L_1(T) = \pi^2 T^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{r(n)^2}{n^{3/2}} - \pi^2 T^{\frac{1}{2}} \sum_{n > T} \frac{r(n)^2}{n^{3/2}} + O(T^{\epsilon}).$$

Consequently, from (6),

$$L_1(T) = \pi^2 T^{\frac{1}{2}} \sum_{n=1}^{\infty} \frac{r(n)^2}{n^{3/2}} + O(T^{\epsilon}). \tag{16}$$

Similarly, from (15) and (11),

$$L_2(T) = O\left(T^{3/2} \sum_{n > T} \frac{r(n)^2}{n^{5/2}}\right) = O(T^{3/2} \sum_{n > T} n^{\epsilon - 5/2}),$$

by (6), where $\epsilon > 0$ is arbitrary; so that

$$L_2(T) = O(T^{\epsilon}). \tag{17}$$

Also, from (15) and (9),

$$L_3(T) = O(\sum_{\substack{n < m \\ m \leqq T}} \frac{r(m) r(n)}{m^{5/4} n^{5/4}} \frac{m^{\frac{1}{2}} n^{\frac{1}{2}}}{m^{\frac{1}{2}} - n^{\frac{1}{2}}} \log T).$$

Hence from $(m^{\frac{1}{2}} - n^{\frac{1}{2}})^{-1} = (m^{\frac{1}{2}} + n^{\frac{1}{2}})/(m - n)$,

$$L_3(T) = O(\sum_{m=1}^{[T]} \frac{m^{\epsilon}}{m^{3/4}} \sum_{n=1}^{m-1} \frac{n^{\epsilon}}{n^{3/4}} \frac{m^{\frac{1}{2}} + n^{\frac{1}{2}}}{m - n} \log T),$$

by (9), where $\epsilon > 0$ is arbitrary. Thus

$$L_3(T) = O(\sum_{m=1}^{[T]} m^{\epsilon - 1/4} \sum_{n=1}^{m-1} \frac{n^{\epsilon - 3/4}}{m - n} \log T).$$

Since this implies that

$$L_3(T) = O(\sum_{m=1}^{[T]} \frac{\log m}{m^{1-2\epsilon}} \log T),$$

where $\epsilon > 0$ is arbitrarily fixed, it follows that

$$L_3(T) = O(T^{\epsilon}) \tag{18}$$

for every $\epsilon > 0$.

Finally, from (15) and (9),

$$L_4(T) = O\left(\sum_{\substack{n < m \\ m > T}} \frac{r(m) r(n)}{m^{5/4} n^{5/4}} \frac{T + T^{\frac{1}{2}} m^{\frac{1}{2}}}{m^{\frac{1}{2}} - n^{\frac{1}{2}}}\right).$$

Hence, from (6),

$$L_4(T) = O\Big(\sum_{m=[T]}^{\infty} \sum_{n=1}^{m-1} \frac{m^{\epsilon} n^{\epsilon}}{m^{5/4} n^{5/4}} \frac{T^{\frac{1}{2}} + m^{\frac{1}{2}}}{m^{\frac{1}{2}} - n^{\frac{1}{2}}} \cdot T^{\frac{1}{2}} \Big).$$

Consequently, from $(m^{\frac{1}{2}} - n^{\frac{1}{2}})^{-1} = (m^{\frac{1}{2}} + n^{\frac{1}{2}})/(m-n)$,

$$L_4(T) = O\Big(\sum_{m=[T]}^{\infty} m^{\epsilon-1/4} \sum_{n=1}^{m-1} \frac{n^{\epsilon-5/4}}{m-n} T^{\frac{1}{2}} \Big),$$

since $m^{\frac{1}{2}} + n^{\frac{1}{2}} = O(m^{\frac{1}{2}})$ and $T^{\frac{1}{2}} = O(m^{\frac{1}{2}})$ in view of $n \leqq m-1$ and of $m \geqq [T]$. Since

$$\sum_{n=1}^{m-1} \frac{n^{\epsilon-5/4}}{m-n} = O(m^{\delta-1}) \text{ as } m \to \infty,$$

where $\epsilon > 0$, hence also $\delta = \delta(\epsilon) > 0$, is arbitrarily small, it follows that

$$L_4(T) = O\Big(\sum_{m=[T]}^{\infty} m^{\epsilon-5/4} T^{\frac{1}{2}} \Big)$$

for every fixed $\epsilon > 0$. Consequently,

$$(19) \qquad L_4(T) = O(T^{\epsilon-1/4} T^{\frac{1}{2}}) = O(T^{\epsilon+1/4}).$$

On substituting (16), (17), (18), (19) into (15), one sees from (14) that

$$(20) \qquad \int_1^T \frac{S(t)^2}{t} dt = \frac{T^{\frac{1}{2}}}{\pi^2} \sum_{n=1}^{\infty} \frac{r(n)^2}{n^{3/2}} + O(T^{\epsilon+1/4})$$

for every $\epsilon > 0$. On the other hand, from (13) and (12),

$$\dot{R}(t) - \pi t = S(t) + O(t^{\epsilon}).$$

Accordingly,

$$S(t) = Q(t^{\frac{1}{2}}) t^{1/4} + O(t^{\epsilon}),$$

by (2). Consequently, from (20),

$$\int_1^T \frac{Q(t^{\frac{1}{2}})^2}{t^{\frac{1}{2}}} dt = \frac{T^{\frac{1}{2}}}{\pi^2} \sum_{n=1}^{\infty} \frac{r(n)^2}{n^{3/2}} + O(T^{\epsilon+1/4}).$$

Hence, if $t^{\frac{1}{2}}$ is replaced by t,

$$\int_1^T Q(t)^2 dt = \frac{T}{2\pi^2} \sum_{n=1}^{\infty} \frac{r(n)^2}{n^{3/2}} + O(T^{\epsilon+\frac{1}{2}}).$$

This completes the proof of (5).

In order to complete the proof of (3), the existence of the Fourier averages of $Q(t)$ and their representation (4) will now be established.

To this end, choose a real number λ, multiply the identity (1) by $e^{i\lambda t}$, and integrate the product between $t = 1$ and $t = T$, where $T > 1$ is arbitrary.

The integration can be carried out term-by-term, since the series (1) is boundedly convergent. Accordingly,

$$\int_1^T e^{i\lambda t} Q(t)\,dt = \frac{1}{\pi} \sum_{n=1}^{\infty} \frac{r(n)}{n^{3/4}} F_n(T;\lambda) + O(T^{\frac{1}{2}}), \tag{21}$$

where $F_n(T;\lambda)$ is an abbreviation for

$$F_n(T;\lambda) = \int_1^T e^{i\lambda t} \sin\,(2\pi n^{\frac{1}{2}} t - \tfrac{1}{4}\pi)\,dt. \tag{22}$$

On writing $\sin\alpha$ in the form $-\frac{1}{2}i(e^{i\alpha} - e^{-i\alpha})$, one easily verifies from (22) that there exists for every λ an $A = A(\lambda)$ such that

$$|F_n(T;\lambda)| < \frac{A(\lambda)}{|\lambda + n^{\frac{1}{2}}|} + \frac{A(\lambda)}{|\lambda - n^{\frac{1}{2}}|} \quad \text{if } \lambda \neq \pm 2\pi n^{\frac{1}{2}} \tag{23}$$

and

$$\left| F_n(T;\lambda) \pm \frac{T}{2i} \exp\,(\pm i\tfrac{1}{4}\pi)\right| < \frac{A(\lambda)}{|\lambda \pm n^{\frac{1}{2}}|} \quad \text{if } \lambda = \pm 2\pi n^{\frac{1}{2}}. \tag{24}$$

Let λ be fixed and let $T \to \infty$.

Suppose first that λ is so chosen that λ^2 is not a multiple of $4\pi^2$. Then (23) is applicable to every n. Hence, (21) implies that

$$\left| \int_1^T e^{i\lambda t} Q(t)\,dt \right| < C(\lambda) \sum_{n=1}^{\infty} \frac{r(n) n^{-\frac{1}{2}}}{n^{3/4}} + O(T^{\frac{1}{2}}) \quad \text{if } \lambda \neq \pm 2\pi n_0^{\frac{1}{2}}, \tag{25}$$

where $n_0 = 1, 2, 3, \cdots$; it being understood that $C = C(\lambda)$ is a function of λ alone.

If, on the other hand, λ is so chosen that either $\lambda = 2\pi n_0^{\frac{1}{2}}$ or $\lambda = -2\pi n_0^{\frac{1}{2}}$, where n_0 is a fixed positive integer, then, on applying (24) to the n_0-th term of (21), and (23) to all remaining terms of (21), one sees that (25) must be replaced by

$$\begin{aligned} &\left| \int_1^T e^{i\lambda t} Q(t)\,dt \pm \frac{T}{2i} \exp\,(\pm i\tfrac{1}{4}\pi) \cdot \frac{1}{\pi} \frac{r(n_0)}{n_0^{3/4}} \right| \\ &\qquad < C(\lambda) \sum_{n=1}^{\infty} \frac{r(n) n^{-\frac{1}{2}}}{n^{3/4}} + O(T^{\frac{1}{2}}) \quad \text{if } \lambda = \pm 2\pi n_0^{\frac{1}{2}}. \end{aligned} \tag{26}$$

Since (6) implies that (25) can be written as

$$\int_1^T e^{i\lambda t} Q(t)\,dt = O(T^{\frac{1}{2}}), \qquad \lambda \neq \pm 2\pi n_0^{\frac{1}{2}}$$

and (26) as

$$\int_1^T e^{i\lambda t} Q(t)\,dt = \pm \frac{T}{2\pi i} \frac{r(n_0)}{n_0^{3/4}} \exp\,(\pm i\tfrac{1}{4}\pi) + O(T^{\frac{1}{2}}), \qquad \lambda = \pm 2\pi n_0^{\frac{1}{2}},$$

the proof of (4) is complete.

Appendix on Waring's Problem.[5]

Let $r_k^{(s)}(n)$ denote the number of ways a positive integer n can be represented as the sum of k-th powers of exactly s positive integers, where $k = 2, 3, 4, \cdots$. Let

$$\text{(I)} \qquad s > (k-2)2^{k-1} + 4; \qquad (k \geqq 2),$$

and let $S_k^{(s)}(n)$ denote the corresponding singular series of Hardy and Littlewood, that is,

$$\text{(II)} \qquad S_k^{(s)}(n) = \sum_{q=1}^{\infty} \sum_{\substack{p (\mathrm{mod}\, q) \\ (p,q)=1}} \left(\frac{1}{q} \sum_{h=0}^{q-1} e_q(h^k p) \right)^s e_q(-np),$$

where $e_q(x) = \exp \dfrac{2\pi i x}{q}$. Then[6]

$$\text{(III)} \qquad r_k^{(s)}(n) \sim C_k^{(s)} n^{-1+s/k} S_k^{(s)}(n) \quad \text{as } n \to \infty,$$

where $C_k^{(s)}$ denotes the positive constant $\Gamma(1 + 1/k)^s / \Gamma(s/k)$.

If s and k are fixed and satisfy (I), the singular series (II) is uniformly convergent with respect to n. In fact, if k is fixed, then[7]

$$\operatorname*{Max}_{p} \left| \sum_{h=0}^{q-1} e_q(h^k p) \right| = O(q^{\epsilon + 1 - 2^{1-k}}) \quad \text{as } q \to \infty$$

holds for every $\epsilon > 0$. Hence, the series (II) is majorized, for every n, by a constant multiple of

$$\sum_{q=1}^{\infty} \left(\frac{1}{q}\, q^{\epsilon + 1 - 2^{1-k}} \right)^s \sum_{\substack{p (\mathrm{mod}\, q) \\ (p,q)=1}} 1 \leqq \sum_{q=1}^{\infty} (q^{\epsilon - 2^{1-k}})^s q,$$

the inner sum on the left being Euler's $\phi(q)$, which is at most q. But the series on the right is convergent if $(\epsilon - 2^{1-k})s + 1 < -1$. Finally, this inequality is satisfied for a sufficiently small $\epsilon > 0$ whenever s satisfies the condition $s > 2^k$; a condition which is implied by (I).

---

[5] This Appendix was received April 1, 1941.

[6] G. H. Hardy and J. E. Littlewood, "Some problems of Partitio Numerorum, I: A new solution of Waring's problem," *Göttinger Nachrichten*, 1920, pp. 33-54. The case $k = 2$ of squares is there excluded, although the results hold for $k = 2$ also; cf. G. H. Hardy, "On the representation of a number as the sum of any number of squares, and in particular of five," *Transactions of the American Mathematical Society*, vol. 21 (1920), pp. 255-284. As to the inequality (I), cf. G. H. Hardy and J. E. Littlewood, "The singular series in Waring's problem and the value of the number $G(k)$," *Mathematische Zeitschrift*, vol. 12 (1922), pp. 161-188.

[7] G. H. Hardy and J. E. Littlewood, *loc. cit.*[6], last formula line of § 5.2 and first formula line of § 2.1 (where $k > 2$; if $k = 2$, the estimate of the Gaussian sums shows that $\epsilon > 0$ can be replaced by $\epsilon = 0$).

Since a uniformly convergent trigonometric series in n always represents a uniformly almost periodic function,[8] having the trigonometric series as Fourier series, it follows that the function (II) of n is uniformly almost periodic, and that (II) remains valid if the sign of equality is replaced by the sign of Fourier equivalence.

Since (I) implies that $s > k$, the exponent of n on the right of (III) is positive, and so

$$\text{(IV)} \qquad n^{1-s/k} r_k^{(s)}(n)$$

differs from $C_k^{(s)} S_k^{(s)}(n)$ by an additive term which tends to zero as $n \to \infty$. It follows therefore from the uniform almost periodicity of (II), that the function (IV) of n is almost periodic, not only in the sense of Besicovitch (B^m), but even in the sense of Weyl (W^m), for arbitrarily large values of the index m; and that the Fourier series (W^∞) of (IV) differs from the uniformly convergent trigonometric series (II) only by the constant factor $C_k^{(s)} > 0$.

It follows also that (IV) is uniformly almost periodic for exactly those pairs s, k for which the asymptotic formula (III) becomes an identity in n. (According to Hardy,[6] this is the case for $k = 2$ if $s \leqq 8$).

These facts, when particularized to the case $k = 2$ of squares, imply the results obtained by Kac[9] on the basis of a heavy analysis,[10] which is now seen to be superfluous (even if $k = 2$ is replaced by $k \geqq 2$). Correspondingly, it follows that the answer to the question raised *loc. cit.*[9] regarding the almost periodicity (B^2) of the case $k = 2$ for $s > 8$ is in the affirmative, since one has almost periodicity (B^∞) and, as a matter of fact, even more than almost periodicity (W^∞), where, in addition, the restriction $k = 2$ can be omitted.

THE JOHNS HOPKINS UNIVERSITY.

---

[8] That is, a sequence which is almost periodic in the sense of Bohr.

[9] M. Kac, "Almost periodicity and the representation of integers as sums of squares," *American Journal of Mathematics*, vol. 62 (1940), 122-126.

[10] *Ibid.*, pp. 123-124.

ON RIEMANN'S FRAGMENT CONCERNING ELLIPTIC MODULAR FUNCTIONS.*

By AUREL WINTNER.

Riemann's posthumous fragment,[1] based on § 40 of Jacobi's Fundamenta Nova,[2] consists of two parts, I and II, which, possibly, were not written consecutively (all that can be taken for granted is that II was written after I). According to Dedekind's comments,[3] the manuscript of II, which contains practically no text, but only formulae, was hardly readable. Dedekind[3] succeeded in explaining and proving the statements of II on the basis of the connection between Gaussian sums and the linear transformation formulae of the elliptic theta functions (Jacobi, Hermite). After Dedekind, further presentations of Riemann's results in II were given by Smith, Hardy, and others.[4]

However, everything in II deals only with the problem of the "heaviest singularities," that is, with the behavior of the function at the rational points of the natural boundary; in other words, with asymptotic formulae valid when the independent variable tends to a root of unity along a radius (or, more generally, along a Stolz path) within the unit circle. On the other hand, I deals with formal expansions of the boundary function (if any!) on the whole of the unit circle; and all that appears to be available in the literature concerning I are certain comments of Smith.[4] The object of the present note is to legitimize Riemann's formal calculations in I, and to show that the problem can be treated in a simple manner on the basis of Lebesgue's theory of integration and of Fourier series.

Dedekind[3] mentions at the beginning of his comments on Riemann's fragments, that the manuscript of I dates from the time (1852) when Riemann was trying to find examples of functions which are represented by trigono-

* Received January 22, 1941.

[1] B. Riemann, *Mathematische Werke*, 2nd edition, pp. 455-461.

[2] C. G. J. Jacobi, *Gesammelte Werke*, vol. 1, pp. 159-164.

[3] *Loc. cit.*[1], pp. 466-478 [= R. Dedekind, *Gesammelte mathematische Werke*, vol. 1, pp. 159-172].

[4] R. Dedekind, *Gesammelte mathematische Werke*, vol. 1, pp. 174-201; H. J. S. Smith, *Collected Papers*, vol. 2, pp. 312-320; G. H. Hardy, "Note on the limiting values of the elliptic modular functions," *Quarterly Journal of Mathematics*, vol. 34 (1903), pp. 76-86; cf. also H. Rademacher, "Zur Theorie der Modulfunktionen," *Journal für reine und angewandte Mathematik*, vol. 167 (1931), pp. 312-336.

metric series but are discontinuous [or, rather, unbounded] everywhere dense. Obviously, Dedekind has in mind the last paragraphs in § 13 of Riemann's paper on trigonometric series.[5] One of the series mentioned there was recently treated[6] by the present methods of the theory of real functions. It turns out that the method applied there to that series can be so adapted as to lead to a corresponding treatment and legalization of the formal trigonometric series found by Riemann in I. It is perhaps of historical interest that Lebesgue's integration theory applies without difficulty precisely in the cases of those transcendents by means of which Riemann himself[5] was illustrating the limitations to which his integration theory subjects the theory of Fourier series.

Actually, the approach to be followed applies not only to elliptic modular functions, but to a large class of expansions; expansions which result from more or less arbitrary Lambert series by the same procedure as the one to which Riemann subjects, in I, the particular Lambert series of the theory of elliptic modular functions given in § 40 of Jacobi's Fundamenta. Incidentally, the particular series mentioned before results precisely if the underlying Lambert series is the one which was considered by Lambert himself,[7] that is, the Lambert series in which every coefficient is 1.

Riemann starts with the expansions 1) — 7) of Jacobi's § 40. Since Riemann's calculations are, at least in principle, the same in all seven cases, it will be sufficient to consider one of them; for instance, Jacobi's[2] series 4). This series represents the complete elliptic integral of the first kind and can be written in the form

$$\frac{2K}{\pi} - 1 = 4 \sum_{n=1}^{\infty} \frac{(-1)^{n+1} q^{2n+1}}{1 - q^{2n+1}},$$

where every value q within the circle $|q| < 1$ is allowed. Riemann divides this odd series by q and integrates the result between $q = 0$ and an arbitrary $q = re^{i\theta}$, where $r = |q| < 1$. This leads to his formula (51).

Thus far everything is legitimate. But now Riemann replaces $r < 1$ by $r = 1$ and thus obtains his final result, (65), in a formal way. Apparently, he first wanted to justify this by using Jacobi's[2] power series 30), that is, the series

$$\frac{2K}{\pi} - 1 = 4 \Sigma\, \psi(n) q^{2^l (4m-1)^2 n},$$

where l, m, n run through all non-negative integers and $\psi(n)$ denotes the

[5] B. Riemann, *loc. cit.*[1], pp. 262-264.

[6] A. Wintner, "On a trigonometrical series of Riemann," *American Journal of Mathematics*, vol. 59 (1937), pp. 629-634.

[7] J. H. Lambert, *Anlage zur Architektonik* . . . , vol. 2 (1771), § 875.

number ($\geqq 0$) of those divisors of n which are of the form $4k+1$. That Riemann's original plan involved this power series, is clear, on the one hand, from his formula (58), and, on the other hand, from the fact that after (58), and then apparently in a separate portion of the fragment,[8] he quotes Abel's continuity theorem. However, the manuscript breaks off both times with a "facile deducitur . . .".

Obviously, this can be explained only by the assumption that, after reaching this critical point in the manuscript, Riemann realized the inapplicability of Abel's continuity theorem. In fact, not only does the latter theorem presuppose the convergence of the series for $r=1$, but this question of convergence is one of the main issues in the problem at hand (in this regard, cf. also the comments of Smith, *loc. cit.*[4]).

Actually, the problem of convergence for $r=1$ and for a specific θ, where $q=re^{i\theta}$, is not the only question resulting from Riemann's formal calculations. In fact, one must ask whether or not the function, which is regular within the circle $|q|<1$, tends to a measurable boundary function as $r\rightarrow 1$; and, if the answer is affirmative, whether or not this boundary function is L-integrable, so that Riemann's formal *trigonometric series* actually is the *Fourier series* (L) of the boundary function (in this regard, cf. a well-known theorem of Privaloff).

It turns out that, on the basis of the development of the theory of real functions since Riemann, one can obtain satisfactory answers to all these questions.

Use will be made of the fact that, if not only the Fischer-Riesz condition

$$(1) \qquad \Sigma\,|\,c_n\,|^2 < \infty$$

for a Fourier series $\Sigma c_n e^{in\theta}$ of a function $f(\theta)$ of class (L^2) is satisfied, but there exists an $\epsilon > 0$ such that

$$(2) \qquad \Sigma\,|\,n\,|^{\epsilon}|\,c_n\,|^2 < \infty,$$

then $\Sigma c_n e^{in\theta}$ is convergent almost everywhere (and represents, by Lebesgue's theorem on $(C,1)$-summability, exactly the value of the function almost everywhere).[9] Needless to say, (2) is neither sufficient nor necessary in order that the sufficient condition

$$(3) \qquad \Sigma\,|\,c_n\,|^{p/(p-1)} < \infty, \qquad (p \geqq 2),$$

for an f of class (L^p) be satisfied by some $p>2$.

---

[8] B. Riemann, *loc. cit.*[1], pp. 460-461.

[9] Actually, it is known that $|\,n\,|^{\epsilon}$ in (2) can be replaced by $\log|\,n\,|$; cf. A. Kolmogoroff and G. Seliverstoff, "Sur la convergence des séries de Fourier," *Rendiconti della Reale Accademia Nazionale dei Lincei*, Ser. 6, vol. 3 (1926), pp. 307-310.

Let $a_1, a_2, \cdots$ be a sequence of constants for which the power series $a_1 z + a_2 z^2 + \cdots$ is convergent in the circle $|z| < 1$, i. e., for which

$$\overline{\lim_{n \to \infty}} \, |a_n|^{1/n} \leqq 1. \tag{4}$$

Then the Lambert series

$$f(z) = \sum_{n=1}^{\infty} a_n \frac{z^n}{1 - z^n} \qquad (|z| < 1) \tag{5}$$

not only is convergent in the circle $|z| < 1$ and represents there a regular function $f(z)$ but the power series of this function can be obtained by formal rearrangement of (5); so that

$$f(z) = \sum_{n=1}^{\infty} b_n z^n \qquad (|z| < 1), \tag{6}$$

where

$$b_n = \sum_{d|n} a_d; \tag{7}$$

it being understood that the summation index d in (7) runs through all divisors of n.

Put $z = re^{i\theta}$, where $r < 1$, and integrate $f(re^{i\theta})/r$ along a radius $\theta =$ const. between $r = 0$ and a fixed $r < 1$. This leads, according to (5), to the function

$$F = \int_0^r \frac{f(se^{i\theta})}{s} ds = \sum_{n=1}^{\infty} a_n e^{in\theta} \int_0^r \frac{s^{n-1} ds}{1 - s^n e^{in\theta}} \tag{8}$$

of r and θ, which is a regular function of $z = re^{i\theta}$; in fact, from (6),

$$F = \sum_{n=1}^{\infty} b_n e^{in\theta} \int_0^r s^{n-1} ds = \sum_{n=1}^{\infty} b_n e^{in\theta} r^n / n. \tag{8 bis}$$

Accordingly,

$$F(z) = \sum_{n=1}^{\infty} c_n z^n, \qquad |z| < 1, \tag{9}$$

where

$$c_n = \frac{1}{n} \sum_{d|n} a_d, \tag{10}$$

by (7).

Obviously, the Lambert series (5) becomes, up to the additive constant 1, Jacobi's above-mentioned expansion for $2K/\pi$, if one chooses

$$a_{2n} = 0, \qquad a_{2n+1} = 4(-1)^{n+1}.$$

Hence, Riemann's procedure, described above, results if one replaces $e^{in\theta}$ in (8 bis) by $e^{i(n+1)\theta}$. The modification of Riemann's procedure by the trivial factor $e^{i\theta}$ appears to be convenient in order to make $F = F(r, \theta)$ a function of $re^{i\theta}$.

Suppose now that the coefficients a_n of (5) are subject, instead of (4), to the sharper restriction that

(11) $$a_n = O(n^{\frac{1}{2}-\epsilon})$$

for some $\epsilon > 0$ (in the above example, $\epsilon = \frac{1}{2}$ is admissible). According to (10) and (11),

$$c_n = O(n^{-\frac{1}{2}-\epsilon})\tau(n),$$

where

(12) $$\tau(n) = \sum_{d|n} 1$$

denotes the number of the divisors of n. Since [10]

(13) $$\tau(n) = O(n^{\delta}) \text{ for every } \delta > 0,$$

it follows that

(14) $$c_n = O(n^{-\frac{1}{2}-\eta}) \text{ for some } \eta > 0.$$

Hence, not only (1) but even (2) is satisfied.

Now let $0 < s < r < 1$. Then, from (9),

$$\int_0^{2\pi} | F(re^{i\theta}) - F(se^{i\theta}) |^2 d\theta = 2\pi \sum_{n=1}^{\infty} | c_n |^2 (r^{2n} - s^{2n}),$$

by Parseval's relation. Hence, from (1),

$$\int_0^{2\pi} | F(re^{i\theta}) - F(se^{i\theta}) |^2 d\theta \to 0 \text{ as } s \to 1, \ r \to 1.$$

Since the Hilbert space is complete, it follows that there exists a function $F(e^{i\theta})$, $0 \leqq \theta < 2\pi$, of class (L^2) such that

(15) $$\int_0^{2\pi} | F(re^{i\theta}) - F(e^{i\theta}) |^2 d\theta \to 0 \text{ as } r \to 1.$$

Since this means strong convergence, and implies therefore weak convergence, the relation

$$\int_0^{2\pi} F(re^{i\theta}) e^{in\theta} d\theta \to \int_0^{2\pi} F(e^{i\theta}) e^{in\theta} d\theta, \text{ as } r \to 1,$$

follows for every n; so that, from (9),

(16) $$F(e^{i\theta}) \sim \sum_{n=1}^{\infty} c_n e^{in\theta}.$$

---

[10] For sharper and more general results, cf. P. Hartman and R. Kershner, "On upper limit relations for number theoretical functions," *American Journal of Mathematics*, vol. 62 (1940), pp. 780-786.

Furthermore, since (2) is satisfied,

$$F(e^{i\theta}) = \sum_{n=1}^{\infty} c_n e^{in\theta} \text{ almost everywhere.} \tag{17}$$

This implies, by Abel's continuity theorem, that not only (15) holds but also

$$F(re^{i\theta}) \to F(e^{i\theta}) \text{ almost everywhere, as } r \to 1. \tag{18}$$

In addition, the radial approach may be replaced by Stolz paths for almost all θ.

If (11) is replaced by the more general condition that

$$a_n = O(n^{\lambda-\epsilon}) \text{ for a fixed positive } \lambda < \tfrac{1}{2} \tag{19}$$

and for some $\epsilon > 0$, then (10) and (13) imply that (14) may be replaced by

$$c_n = O(n^{\lambda-1-\eta}) \text{ for some } \eta > 0; \tag{20}$$

so that the sufficient condition (3) for a boundary function $F(e^{i\theta})$ of class (L^p) is satisfied by

$$p = 1/\lambda. \tag{21}$$

If, in particular,

$$a_n = O(n^{\delta}) \text{ for every } \delta > 0, \tag{22}$$

then λ in (19) can be chosen arbitrarily small, and so (21) shows that $F(e^{i\theta})$ is of class (L^∞) [that is, of class (L^p) for every p].

Condition (22) is, of course, satisfied if the a_n are chosen as in the above-mentioned example of $2K/\pi$, and, more generally, if $a_n = O(1)$; for instance, if $a_n = 1$ for every n. In the latter case, (5), (6), (7) and (12) supply Lambert's own expansion [7]

$$f(z) = \sum_{n=1}^{\infty} \frac{z^n}{1 - z^n} = \sum_{n=1}^{\infty} \tau(n) z^n, \qquad |z| < 1,$$

while (9) reduces, in view of (10), to

$$F(z) = \sum_{n=1}^{\infty} \frac{\tau(n)}{n} z^n, \qquad |z| < 1.$$

Consequently, there exists a boundary function $F(e^{i\theta})$ of class (L^∞), having the Fourier series

$$F(e^{i\theta}) \sim \sum_{n=1}^{\infty} \frac{\tau(n)}{n} e^{in\theta}$$

which is convergent almost everywhere. The imaginary part of this series is precisely the expansion treated before [11]; an expansion from which the series

[11] A. Wintner, *loc. cit.*[6]

$$\sum_{n=1}^{\infty} \frac{1}{n} \Big(\sum_{d|n} (-1)^d \Big) \sin n\theta,$$

considered by Riemann in his paper on trigonometric series,[12] follows by writing 2θ for θ and subtracting the result from the expansion itself.

If $a_n = 1$ is replaced by $a_n = n^\alpha$, where $0 < \alpha < \frac{1}{2}$, then (19), (20), (21) and (2) show that there exists a boundary function $F(e^{i\theta})$ which is of class (L^p) for every $p < 1/\alpha$ and possesses a Fourier series which is convergent almost everywhere.[13]

By the theory of the elliptic modular functions, all expansions considered by Jacobi[2] in his § 40 must be identical consequences of the corresponding expansion of the (principal) logarithm of the fundamental invariant

$$\Delta = z \prod_{n=1}^{\infty} (1 - z^n)^{24}, \qquad (|z| < 1);$$

that is, of the expansion

$$-\frac{1}{24} \log \frac{\Delta}{z} = -\sum_{n=1}^{\infty} \log (1 - z^n) \equiv \sum_{n=1}^{\infty} \frac{1}{n} \frac{z^n}{1 - z^n}.$$

Thus, $a_n = 1/n$ in (5). Hence, (22) is certainly satisfied, and so there exists, by (16), (17) and (10), a boundary function of class (L^∞), possessing the Fourier series

$$F(e^{i\theta}) \sim \sum_{n=1}^{\infty} \Big(\frac{1}{n} \sum_{d|n} \frac{1}{d} \Big) e^{in\theta}$$

which is convergent almost everywhere.

It is clear that the facts valid under the respective assumptions (4), (11), (19), (21) hold also when (5) is replaced by

(5 bis) $$f(z) = \sum_{n=1}^{\infty} a_n \frac{z^n}{1 + z^n}.$$

In fact, all that is necessary is to replace (7) by

(7 bis) $$b_n = -\sum_{d|n} (-1)^d a_{n/d}$$

and, correspondingly, (10) by

(10 bis) $$c_n = -\frac{1}{n} \sum_{d|n} (-1)^d a_{n/d}.$$

THE JOHNS HOPKINS UNIVERSITY.

---

[12] B. Riemann, *loc. cit.*[8]

[13] Cf. P. Hartman and A. Wintner, "On certain Fourier series involving sums of divisors," *Travaux de l'Institut Mathématique de Tbilissi*, vol. 3 (1938), pp. 114-118.

A THEOREM AND AN INEQUALITY REFERRING TO RECTIFIABLE CURVES.*

By L. A. Santaló.

Introduction. The principal purpose of this paper is the demonstration of the following theorem:

Given a rectifiable curve of length L in the euclidean space E_m and calling n the number of common points of this with the surface of a sphere of centre $P(x_1, x_2, \cdots, x_m)$ and radius R, we find

$$\int n\, dP = 2LV_{m-1}(R)$$

where $dP = dx_1\, dx_2 \cdots dx_m$ and $V_{m-1}(R)$ represents the volume of the $(m-1)$-dimensional sphere. The integration is extended to all the points P for which $n \neq 0$.

For example: for the plane $(m=2)$ if x_1, x_2 are the coordinates of the centre of a circle of radius R we obtain

$$\int n\, dx_1\, dx_2 = 4LR,$$

For space $(m=3)$

$$\int n\, dx_1\, dx_2\, dx_3 = 2L\pi R^2.$$

In all cases n is an integral-valued function of x_i.

In § 4 we give some applications of this formula. For the case of the plane the formula may be considered as a particular case of the so called formula of Poincaré of integral geometry and it has been demonstrated more generally by Maak [1].[1] When $m = 3$ and under the assumption that the curve possesses a continuously turning tangent (or is composed of a finite number of arcs of this nature) we have obtained this result in another paper [2].

§ 1. Notations.

Let us represent by $V_m(R)$ the volume of the m-dimensional sphere of radius R which is, as is known,

$$V_m(R) = \frac{(\pi R^2)^{m/2}}{\Gamma(m/2+1)}. \tag{1}$$

The volume of a spherical segment of this sphere which has a semiangle

* Received August 20, 1940.

[1] Numbers in brackets refer to the bibliography at the end of the paper.

in the center equal to α may be represented by $S_m(\alpha, R)$ and may be calculated in the following form. On intersecting the sphere by an hyperplane at a distance x from the center, the section is an $(m-1)$-dimensional sphere whose volume is $V_{m-1}(\sqrt{R^2-x^2})$ and the volume of the spherical segment of one base distant from the center $h = R\cos\alpha$ will be

$$(2) \qquad S_m(\alpha, R) = \int_h^R V_{m-1}(\sqrt{R^2-x^2})\,dx.$$

Placing $x = R\cos\theta$ and using (1)

$$(3) \qquad S_m(\alpha, R) = \frac{\pi^{(m-1)/2}R^m}{\Gamma\left(\frac{m+1}{2}\right)}\int_0^\alpha \sin^m\theta\,d\theta = V_{m-1}(R)\cdot R\cdot\int_0^\alpha \sin^m\theta\,d\theta.$$

On evaluating the last integral by integration by parts and writing $\beta = \frac{\pi}{2} - \alpha$ we obtain:

For m even

$$(4) \qquad S_m(\alpha, R) = RV_{m-1}\left[\frac{(m-1)(m-3)\cdots 3.1}{m(m-2)\cdots 4.2}\frac{\pi}{2} - \sin\beta\right.$$
$$\left(\frac{1}{m}\cos^{m-1}\beta + \frac{(m-1)}{m(m-2)}\cos^{m-3}\beta + \cdots + \frac{(m-1)(m-3)\cdots 5.3}{m(m-2)\cdots 4.2}\cos\beta\right)$$
$$\left. - \frac{(m-1)(m-3)\cdots 5.3}{m(m-2)\cdots 4.2}\beta\right];$$

and for m odd

$$(5) \qquad S_m(\alpha, R) = RV_{m-1}\left[\frac{(m-1)(m-3)\cdots 4.2}{m(m-2)\cdots 3.1} - \sin\beta\left(\frac{1}{m}\cos^{m-1}\beta\right.\right.$$
$$\left.\left. + \frac{(m-1)}{m(m-2)}\cos^{m-3}\beta + \cdots + \frac{(m-1)(m-3)\cdots 4.2}{m(m-2)\cdots 3.1}\right)\right].$$

For $m = 2$ the formula given above should be replaced by

$$(6) \qquad S_2(\alpha, R) = V_1(R)R\left[\frac{1}{2}\frac{\pi}{2} - \frac{1}{2}(\sin\beta\cos\beta + \beta)\right].$$

§ 2. Case of a polygonal line.

1. Measure of a set of spheres. Let E_m represent the euclidean space of dimension m. Let us call the measure of a set of spheres of radius R, the measure in E_m of the set of their centres.

Consequences. a). Given a finite number or an enumerable infinity of points, the measure of the spheres, whose surfaces contain some of these, is zero.

b). The measure of the spheres tangent to a segment of a straight line is zero.

2. Spheres which have common points with a segment. For the case of the plane (fig. 1), and the result is completely general for a space of any number of dimensions, the measure M_0 of the set of spheres of radius R which contains totally in its interior a segment of length l $(l \leq 2R)$, is the part indicated by points, or twice the volume of the spherical segment of a sphere of radius R and semiangle $\alpha = \text{arc}\cos\frac{l}{2R}$; in accordance with the notation (2)

$$M_0 = 2S_m(\alpha, R). \tag{7}$$

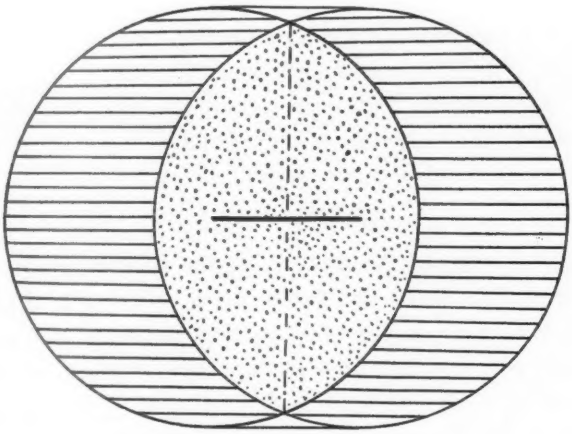

Fig. 1.

The measure M_1 of the spheres of radius R which have with the segment only one point of intersection is composed of two "lunules" (see the shaded area in fig. 1 for E_2), the volume of which will be, in E_m

$$M_1 = 2V_m(R) - 4S_m(\alpha, R). \tag{8}$$

The total measure of the m-dimensional spheres of radius R which have common points with a segment of length l is equal to the volume of a cylinder of height l whose base is an $(m-1)$-dimensional sphere of radius R, plus two hemispheres of radius R. This measure is the sum of the measure of those spheres which contain the segment in their interior (M_0), plus the measure of the spheres which have with the segment one only point of intersection (M_1), plus the measure of the spheres which have with the segment two points of intersection (M_2). Therefore

$$M_0 + M_1 + M_2 = V_m(R) + lV_{m-1}(R)\,. \tag{9}$$

By (7), (8) and (9) we obtain

$$M_2 = lV_{m-1} - V_m + 2S_m(\alpha, R) \tag{10}$$

and by (8)

$$M_1 + 2M_2 = 2lV_{m-1}. \tag{11}$$

If $x_1, x_2, \cdots, x_m$ are the coordinates of the centre of the sphere, we find on putting $dP = dx_1\, dx_2 \cdots dx_m$ and representing by n the number of the points of intersection of the sphere with the fixed segment of length l, that (11) is equivalent to

$$\int n\, dP = 2lV_{m-1} \tag{12}$$

where n is an integral-valued function of $x_1, x_2, \cdots, x_m$.

3. Case of a polygonal line. Let us consider a polygonal line the sides of which have lengths l_i, the total length being $\Sigma l_i = L$.

If we add the integrals (12) corresponding to each side and still denote by n the number of common points of the polygonal line and the surface of the sphere of centre P and radius R, we obtain

$$\int n\, dP = 2LV_{m-1}. \tag{13}$$

4. Maximum value of the measure of the spheres whose surface have common points with a closed rectifiable curve. We denote the length of the curve by L and consider an inscribed polygonal line σ_i of length L_i. The measure of the set of spheres whose surfaces have only one point in common with the polygonal line is zero, because this set is composed of spheres tangent to the sides plus the spheres whose surfaces contain some vertices. Representing by $M_t(\sigma_i)$ the total measure of the spheres which have common points with the polygonal line, we have by (13)

$$M_t(\sigma_i) \leq \tfrac{1}{2}\int n\, dP = L_i V_{m-1} \leq LV_{m-1} \tag{14}$$

because, as we have seen, save for a set of positions of measure zero, n is ≥ 2.

When the number of sides of the polygonal line grows indefinitely, and at the same time their lengths tend towards zero, the volume (14) filled by the centres of the spheres whose surfaces have points in common with the polygonal line, tends towards the volume M_t corresponding to the centres of the spheres whose surfaces have some points in common with the curve. Consequently, by (14) we obtain in the limit

$$M_t \leq LV_{m-1}. \tag{15}$$

5. Lemma. We consider a rectifiable curve of lenth L and denote by l_i the length of the sides of an inscribed polygonal line. Summing Σl_i^r extended on every side, where $r = 1 + a$ and $a > 0$, we have

$$\lim \Sigma l_i^r = 0. \tag{16}$$

This limit is considered for any succession of inscribed polygonal lines which tend towards the curve. In fact if we denote by λ_i the side of greatest length, we have

$$\lim \Sigma l_i^r \leq \lim (\lambda_i^a \Sigma l_i) = L \lim \lambda_i^a = 0.$$

As a consequence we observe that, since

$$l_i - \frac{l_i^3}{3!} < \sin l_i < l_i$$

we have

(17) $$\lim \Sigma \sin l_i = L.$$

§ 3. Case of a rectifiable curve.

1. Theorem. If a curve contained in E_m is rectifiable and has length L, and we denote by n the number of common points with the surface of a sphere of radius R and centre P $(x_1, x_2, \cdots, x_m)$ and put $dP = dx_1\, dx_2 \cdots dx_m$, we obtain

(18) $$\int n\, dP = 2LV_{m-1}(R)$$

The radius R may have any value. The integration is extended to all the points for which $n \neq 0$.

Proof. Let us consider two classes of points common to the curve and the surface of the sphere: I. *Intersection points,* when in any neighborhood of these there are points of the curve interior and exterior to the sphere. II. *Contact points,* when in the neighborhood of these the points of the curve are all from the interior or from the exterior of the sphere.

Let us indicate by $n^{(i)}$ the number of the first class for any position of the sphere, and by $n^{(c)}$ the number of the second group. We have $n = n^{(i)} + n^{(c)}$ and both numbers are functions of the coordinates $x_1, x_2, \cdots, x_m$ of the centre of the sphere.

I. We shall first consider the intersection points. Let us take a polygonal line σ_1 inscribed in the curve. For each position of the sphere, we may represent by $n_1^{(i)}$ the number of points of intersection with the polygonal line in which the corresponding side l_i presents one extremity outside and the other inside the sphere, plus the points of intersection which are vertices of the polygonal line and in which this cuts through the spherical surface, the other extremes of these sides which concur in this point remaining in a distinct region of the sphere.

By (8) we obtain

(19) $$\int n_1^{(i)}\, dP = \sum_i (2V_m - 4S_m(\alpha_i, R))$$

in which $S_m(\alpha_i, R)$ has the value (3) with $\alpha_i = \text{arc cos} \dfrac{l_i}{2R}$ and the sum is extended to all the sides of the polygonal line.

On increasing the number of sides of the inscribed polygonal line, in order to unite the extremities of each l_i which has an intersection point or to unite the extremities of two consecutive sides which concur at one point of the surface of the sphere being separate from this, it is always necessary to pass from an interior point to another which is exterior to the sphere. In consequence for a constant position of P, the number $n_1^{(i)}$ does not diminish when the number of sides of the polygonal line increases. That is to say

$$n_1^{(i)} \leq n_2^{(i)} \leq \cdots \leq n_\nu^{(i)} \leq \cdots$$

and when the number of sides increases and their respective lengths tend towards zero, we obtain $\lim\limits_{\nu\to\infty} n_\nu^{(i)} = n^{(i)}$.

In accordance with the fundamental property of the Lebesgue integral, we have

$$\int n^{(i)}\, dP = \lim_{\nu\to\infty} \int n_\nu^{(i)}\, dP = \lim \sum_i (2V_m - 4S_m(\alpha_i, R)). \tag{20}$$

To find this limit we substitute for $S_m(\alpha_i, R)$ its value given by (4), (5) and also set

$$\frac{(m-1)(m-3)\cdots 3.1}{m\,(m-2)\cdots 4.2}\,\frac{\pi}{2}\,RV_{m-1} = \tfrac{1}{2}V_m \qquad \text{for } m \text{ even}$$

and

$$\frac{(m-1)(m-3)\cdots 4.2}{m(m-2)\cdots 3.1}\,RV_{m-1} = \tfrac{1}{2}V_m \qquad \text{for } m \text{ odd.}$$

We obtain, if m is an even number,

$$\begin{aligned} 2V_m - 4S_m(\alpha_i, R) = 4RV_{m-1} \Big[\sin\beta_i \Big(&\frac{1}{m}\cos^{m-1}\beta_i \\ &+ \frac{(m-1)}{m(m-2)}\cos^{m-3}\beta_i + \cdots + \frac{(m-1)(m-3)\cdots 5.3}{m(m-2)\cdots 4.2}\cos\beta_i \Big) \\ &+ \frac{(m-1)(m-3)\cdots 5.3}{m(m-2)\cdots 4.2}\beta_i \Big] \end{aligned}$$

and if m is an odd number

$$\begin{aligned} 2V_m - 4S_m(\alpha_i, R) = 4RV_{m-1}\sin\beta_i \Big(&\frac{1}{m}\cos^{m-1}\beta_i \\ &+ \frac{(m-1)}{m(m-2)}\cos^{m-3}\beta_i + \cdots + \frac{(m-1)(m-3)\cdots 4.2}{m(m-2)\cdots 3.1} \Big) \end{aligned}$$

where

$$\beta_i = \frac{\pi}{2} - \alpha_i = \text{arc sin}\,\frac{l_i}{2R}.$$

On adding these expressions for all sides l_i of the inscribed polygonal line and approximating this polygonal line towards the curve, we have in the limit (considering (16), (17)):

For m even

$$(21) \qquad \lim \sum_i (2V_m - 4S_m(\alpha_i, R))$$
$$= 2LV_{m-1}\left(\frac{1}{m} + \frac{(m-1)}{m(m-2)} + \cdots + 2\frac{(m-1)(m-3)\cdots 5.3}{m(m-2)\cdots 4.2}\right)$$

and for m odd

$$(22) \qquad \lim \sum_i (2V_m - 4S_m(\alpha_i, R))$$
$$= 2LV_{m-1}\left(\frac{1}{m} + \frac{(m-1)}{m(m-2)} + \cdots + \frac{(m-1)(m-3)\cdots 4.2}{m(m-2)\cdots 3.1}\right).$$

But the sums which are in the parentheses are equal to unity in both cases, and so

$$(23) \qquad \int n^{(i)}\, dP = 2LV_{m-1}.$$

II. We must now consider the contact points. Denoting by $n^{(c)}$ the number of these corresponding to a position of the sphere we shall show that

$$(24) \qquad \int n^{(c)}\, dP = 0.$$

Let us consider a side of length l_i of an inscribed polygonal line and the arc of the curve comprehended between its extremes. This arc (of length u_i) plus the side of the polygonal line forms a closed curve of length $l_i + u_i$. Accordingly by (15), the measure M_t of the spheres whose surface has common points with the curve is $M_t \leq (l_i + u_i)V_{m-1}$. Subtracting from this value the measure of the spheres whose surfaces cut the side of the polygonal line at only one point (given by (8)), there results

$$(25) \qquad (l_i + u_i)V_{m-1} - [2V_m - 4S_m(\alpha_i, R)].$$

On summing these values for all sides of the polygonal line we obtain an upper bound to the measure of the spheres whose surface cuts the curve without cutting any side of the polygonal line in one point only (it might occur that the sphere cuts in two points some of the sides). Consider now a sphere with $n^{(c)}$ contact points. From an inscribed polygonal line of sides sufficiently small, this sphere for each point of the $n^{(c)}$ is contained in the sum of the expressions (25). Consequently

$$(26) \qquad \int n^{(c)}\, dP \leq \lim \sum_i [(l_i + u_i)V_{m-1} - 2V_m + 4S_m(\alpha_i, R)]$$

As before, the limit is considered for a succession of inscribed polygonal

lines which tends towards the curve. But $\lim \sum_i (l_i + u_i) V_{m-1} = 2LV_{m-1}$ and having in mind (21) and (22) we obtain

$$\lim \sum_i (2V_m - 4S_m(\alpha_i, R)) = 2LV_{m-1}.$$

(24) is thus demonstrated.

On summing (23) and (24) and taking into account the fact that $n = n^{(i)} + n^{(c)}$ we obtain the formula (18).

2. Reciprocal theorem. *Calling n the number of common points of a Jordan curve with the surface of a sphere of radius R and centre $P(x_1, x_2, \cdots, x_m)$ the curve is rectifiable and its length is equal to $J : 2V_{m-1}$ if the integral*

$$J = \int n \, dP \tag{27}$$

is finite.

Proof. For each position of the sphere we decompose n into $n = n^{(i)} + n^{(c)}$, $n^{(i)}$ being the number of intersection points and $n^{(c)}$ the contact points defined previously. Fixing a sphere of centre $P(x_1, x_2, \cdots, x_m)$ and inscribing a polygonal line in the given curve, and representing by $n_1^{(i)}$ the number of sides of this polygonal line which cut the surface of the sphere in one only point, we have when we inscribe successive polygonal lines, for a constant position of the sphere,

$$n_1^{(i)} \leq n_2^{(i)} \leq \cdots \leq n_\nu^{(i)} \leq \cdots \leq n^{(i)}.$$

Integrating this succession and applying the formula (8) for polygonal lines we obtain

$$\int n_\nu^{(i)} \, dP = \sum_i (2V_m - 4S_m(\alpha_i, R)) \leq \int n^{(i)} \, dP \leq J.$$

Consequently the first sum is bounded as l_i tends towards zero, which shows, on taking into account the values (4) (5), that Σl_i is also bounded. The curve is thus rectifiable and by the direct theorem its length is equal to $J : 2V_{m-1}$.

§ 4. Applications.

1. A definition of length of a curve. The theorem of the previous section permits the following definition of length of a set of points of E_m:

Let us consider a set of points of E_m. Calling n the number of these which belong to the surface of the sphere of radius R and centre $P(x_1, x_2, x_3, \cdots, x_m)$, the length of the set of points is

$$\frac{1}{2V_{m-1}(R)}\int n\,dP \tag{28}$$

where $dP = dx_1, dx_2 \cdots dx_m$ *and the integration is extended over all space or, what is the same, over the positions of* P *for which* $n \neq 0$.

In accordance with the reciprocal theorem of the last section this present definition coincides with the usual definition if the set of points is a Jordan curve.

2. A sufficient condition in order that a curve should be rectifiable. If a curve is situated at finite distance and the maximum number of common points that the curve could have with the surface of a sphere of any radius is limited, the curve is rectifiable.

This is a consequence of the reciprocal theorem. This property is due to Marchaud [3].

3. An inequality for rectifiable curves. THEOREM: *If* C *is a rectifiable curve of length* L *in* E_m *and* V *is the volume filled by the points in space whose distance from any point of the curve is* $\leq R$, *then*

$$V \leq LV_{m-1}(R) + V_m(R). \tag{29}$$

This inequality, for the case of the plane $(m = 2)$, has been obtained by Hornich [4]. The procedure we propose to follow will permit us also to give an interpretation of the "deficit," that is to say, of the difference of the two members of the inequality (29). Let M_i be the measure of the spheres of radius R which have i common points with the curve. In accordance with (18), we have

$$M_1 + 2M_2 + 3M_3 + \cdots = 2LV_{m-1}. \tag{30}$$

If M_0 indicates the measure of the spheres which contain the curve C in their interior, in accordance with the definition of V we obtain also

$$M_0 + M_1 + M_2 + M_3 + \cdots = V. \tag{31}$$

From (30) and (31) we can deduce that

$$M_3 + 2M_4 + \cdots - (2M_0 + M_1) = 2LV_{m-1} - 2V. \tag{32}$$

We consider the segment of length D which unites the extremities of the given curve (if the curve C is closed, $D = 0$). Let us represent by M^*_0 the measure of the spheres of radius R which contain this segment totally in their interior, and by M^*_i $(i = 1, 2)$ the measure of the spheres whose surfaces have i common points with the same segment. According to (9) and (11) we have

$$M^*_0 + M^*_1 + M^*_2 = V_m + DV_{m-1}; \qquad M^*_1 + 2M^*_2 = 2DV_{m-1}$$

so that

$$2M^*_0 + M^*_1 = 2V_m.$$

If the sphere contains the curve C in its interior it will also contain the segment that unites its extremities, and in consequence $M_0 \leq M^*_0$. Also, if the surface of the sphere cuts the curve in only one point, this curve will have one of its extremities in the interior, and the other exterior, and so the segment will cut the sphere also at only one point. That is to say, $M_1 \leq M^*_1$. It follows that

$$2M_0 + M_1 \leq 2M^*_0 + M^*_1 = 2V_m.$$

Applying this inequality to (32), we obtain

$$\tfrac{1}{2}(M_3 + 2M_4 + \cdots) + V \leq LV_{m-1} + V_m \tag{33}$$

which implies (29).

The equality in (33) will be verified, if $M_i = 0$, only by $i \geq 3$ and moreover $M^*_0 = M_0$, $M^*_1 = M_1$. The condition $M_i = 0$ $(i \geq 3)$ carries with it $M^*_1 = M_1$ and in consequence the conditions for equality are:

1°. $M_i = 0$ $(i \geq 3)$. The curve can not be cut by the surface of the sphere in more than two points (with the exception perhaps of a set of positions of zero measure).

2°. $M^*_0 = M_0$. That is to say, every position of the sphere in which it contains in its interior the two extremities of the curve, contains also all the curve.

In particular, if the given curve C is closed, the equality only is valid in the case of reduction to a point.

Rosario (Rép. Argentina).

BIBLIOGRAPHY.

[1]. W. Maak, "Ueber stetige Kurven (Integralgeometrie 27)." *Abhandlungen aus dem Mathematischen Seminar der Hansischen Universität*, vol. 12 (1938), pp. 163-178.

[2]. L. A. Santaló, "Integralgeometrie. Ueber das kinematische Mass im Raum," *Actualités Scientifiques et industrielles No. 357*, Hermann et Cie, editeurs: Paris (1936).

[3]. A. Marchaud, "Sur les continus d'ordre borné," *Acta Mathematica*, vol. 55 (1930), pp. 67-115.

[4]. H. Hornich, "Eine allgemeine Ungleichung für Kurven," *Monatshefte für Mathematik und Physik*, vol. 47 (1939), pp. 432-438.

TENSOR ALGEBRA AND YOUNG'S SYMMETRY OPERATORS.* [1]

By T. L. Wade.

Introduction. In treating the problem of decomposing tensor space into its irreducible components under the full linear group, Weyl [2] has proved that the tensor of each irreducible subspace is generated by a Young symmetrizer operating on the arbitrary tensor of the given tensor space, there being a Young symmetrizer corresponding to each partition of the indices of the given arbitrary tensor. Although, as is well known, the Frobenius-Schur theory of group characters and the substitutional analysis of Young are related, the method of obtaining an explicit algebraic construction of the decomposition of an arbitrary tensor which is commonly given [3] implies the application of Young's symmetry operators as developed in his earlier papers.[4] Even after Young established the relation between his analysis and the Frobenius-Schur theory of group characters,[5] there seemed to be no particular advantage in viewing the decomposition of an arbitrary tensor as the operation on it of operators defined with the aid of the characters of the symmetric group, in lieu of the operation on it of Young's symmetry operators.[6] For while the application of the Young symmetrizers was laborious, the calculation of the characters by the then known methods of Frobenius,[7] Schur,[8] and Burnside [9] was likewise laborious. However, with the development of immanants of matrices by Littlewood and Richardson,[10] and the subsequent use of them [11] in the relatively easier and shorter calculation of characters, and the still more recent work of Murnaghan [12] which made available the calculation of characters for any value of p, the situation is changed. The features of this paper are: (a) the assumption of the availability [13] of the characters, or of expeditious means for their calculation [14]; (b) a specialization by means of characters of Cramlet's general

* Received January 15, 1941.

[1] Presented to the American Mathematical Society, December 30, 1940, and abstracted in the Bulletin of the Society, vol. 47 (1941), p. 32.

[2] [1].

[3] [1], [2].

[4] [3].

[5] [4].

[6] [5], [6], [7], [8].

[7] [9].

[8] [10].

[9] [11].

[10] [12].

[11] [13], [14], [15], [16].

[12] [17], [18], [19].

[13] [12], [15], [16], [19].

[14] [18].

invariant,[15] or absolute numerical, tensor whereby there is associated with each partition of p an absolute numerical tensor; these are called *immanant tensors*; (c) a consideration of the process of the decomposition of an arbitrary tensor with the use of immanant tensors. It seems desirable to divide the paper in three parts: section 1, a brief recounting of Young's symmetry operators; section 2, the introduction of the immanant tensors; and section 3, the application of the immanant tensors to the decomposition process.

1. Young's symmetry operators.[16] Let[17] $a_1, \cdots, a_p$ be p letters arranged in $h \leqq p$ horizontal rows, so that each row has its first letter in the same vertical column, its second letter in a second vertical column, and so on; there being α_1 letters in the first row, α_2 letters in the second row, etc., and finally α_h letters in the last row; where

$$\alpha_1 + \alpha_2 + \cdots + \alpha_h = p,$$
$$\alpha_1 \geqq \alpha_2 \geqq \cdots \geqq \alpha_h$$

Such an arrangement can be represented by the tableau:

$$[\alpha] = [\alpha_1, \alpha_2, \cdots, \alpha_p]: \begin{array}{llll} a_{11} & a_{12} & \cdots & a_{1\alpha_1} \\ a_{21} & a_{22} & \cdots & a_{2\alpha_2} \\ \cdot & \cdot & \cdot & \cdot \\ a_{h1} & a_{h2} & \cdots & a_{h\alpha_h}. \end{array}$$

If $h < p$ we write

$$[\alpha_1, \cdots, \alpha_p] = [\alpha_1, \cdots, \alpha_h] = [\alpha] \quad \text{where} \quad \alpha_{h+1} = \alpha_{h+2} = \alpha_p = 0.$$

Denote the sum of all permutations of the symmetric group on the p letters $a_1, \cdots, a_p$ by $\{a_1 \cdots a_p\}$; Young calls such an expression the "positive symmetric group." Thus in the conventional symbolism of circular permutations,

$$\{a_1a_2a_3\} = 1 + (a_1a_2) + (a_1a_3) + (a_2a_3) + (a_1a_2a_3) + (a_1a_3a_2).$$

From the rows of $[\alpha]$ we can construct the product of the h positive symmetric groups on the letters in the h rows of the tableau. Thus

$$P_\alpha = \{a_{11}a_{12} \cdots a_{1\alpha_1}\}\{a_{21}a_{22} \cdots a_{2\alpha_2}\} \cdots \{a_{h1}a_{h2} \cdots a_{h\alpha_h}\}.$$

Similarly, denote the sum of all the even permutations of the symmetric group on the p letters $a_1, a_2, \cdots, a_p$ minus the sum of all the odd permutations

---

[15] [20].

[16] [3].

[17] The author is indebted to Dr. R. H. Bruck for helpful suggestions concerning this section.

on these letters by $\{a_1a_2 \cdots a_p\}'$; Young calls this the "negative symmetric group." To illustrate,

$$\{a_1a_2a_3\}' = 1 - (a_1a_2) - (a_1a_3) - (a_2a_3) + (a_1a_2a_3) + (a_1a_3a_2).$$

From the columns of $[\alpha]$ we can construct the product of the α_1 negative symmetric groups,

$$N_a = \{a_{11}a_{21} \cdots a_{h1}\}'\{a_{12}a_{22} \cdots a_{h2}\}' \cdots \{a_{1\alpha_1} \cdots\}'.$$

Consider the product of these expressions, P_aN_a, and take the sum of the $p!$ arrangements of the letters in the tableau $[\alpha]$; we write

$$T_{a_1a_2\ldots a_h} = \left(\frac{f_{a_1a_2\ldots a_h}}{p!}\right)^2 \Sigma\, P_aN_a, \tag{1}$$

where

$$f_{a_1a_2\ldots a_h} = p!\,\frac{\prod_{r,s} (\alpha_r - \alpha_s - r + s)}{\prod_r (\alpha_r + h - r)!}. \tag{2}$$

$r < s \leq h$

$r = 1, 2, \ldots, h$

The $T_{a_1a_2\ldots a_h}$'s satisfy the basic substitutional identity

$$1 = \sum T_{a_1a_2\ldots a_h}, \tag{3}$$

and, X being the function to be symmetrized,

$$X = \sum T_{a_1a_2\ldots a_h}X. \tag{4}$$

For $p = 2$, there are two forms of tableaux:

$$a_1a_2\,, \qquad \begin{matrix} a_1 \\ a_2 \end{matrix}\,;$$

and

$$T_2 = \{a_1a_2\}, \qquad T_{1,1} = \{a_1a_2\}'.$$

In this case the identity (3) becomes

$$1 = \frac{1}{2!}\{a_1a_2\} + \frac{1}{2!}\{a_1a_2\}';$$

and this applied to the tensor $F_{i_1i_2}$ gives

$$F_{i_1i_2} = \frac{1}{2!}[F_{i_1i_2} + F_{i_2i_1}] + \frac{1}{2!}[F_{i_1i_2} - F_{i_2i_1}]. \tag{5}$$

For $p = 3$, there are three form of tableaux:

$$a_1a_2a_3, \qquad \begin{matrix} a_1a_2, \\ a_3 \end{matrix} \qquad \begin{matrix} a_1 \\ a_2 \\ a_3 \end{matrix}\,;$$

and

$$T_3 = \frac{1}{3!}\{a_1a_2a_3\};\ T_{2,1} = \left(\frac{2}{3!}\right)^2 \Sigma\{a_1a_2\}\{a_1a_3\}';\ T_{1,1,1} = \frac{1}{3!}\{a_1a_2a_3\}'.$$

Here identity (3) becomes

$$1 = \frac{1}{3!}\{a_1a_2a_3\} + \frac{1}{9}\Sigma\{a_1a_2\}\{a_1a_3\}' + \frac{1}{3!}\{a_1a_2a_3\}',$$

which gives the following result when applied to the tensor $F_{i_1i_2i_3}$:

$$(6)\qquad F_{i_1i_2i_3} = \frac{1}{6}[A+B+C] + \frac{1}{9}[6A-3C] + \frac{1}{6}[A-B+C],$$

where

$$\begin{aligned} A &= F_{i_1i_2i_3}, \\ B &= F_{i_1i_3i_2} + F_{i_2i_1i_3} + F_{i_3i_2i_1}, \\ C &= F_{i_2i_3i_1} + F_{i_3i_1i_2}. \end{aligned}$$

For $p = 4$ there are five forms of tableaux:

$$\begin{array}{ccccc} a_1a_2a_3a_4, & a_1a_2a_3, & a_1a_2, & a_1a_2, & a_1 \\ & a_4 & a_3a_4, & a_3 & a_2 \\ & & & a_4 & a_3 \\ & & & & a_4 \end{array}.$$

In this case the substitutional identity is

$$\begin{aligned} 1 = {} & \frac{1}{4!}\{a_1a_2a_3a_4\} + \frac{1}{64}\Sigma\{a_1a_2a_3\}\{a_1a_4\}' \\ & + \frac{1}{144}\Sigma\{a_1a_2\}\{a_3a_4\}\{a_1a_3\}'\{a_2a_4\}' + \frac{1}{64}\Sigma\{a_1a_2\}\{a_1a_3a_4\}' + \frac{1}{4!}\{a_1a_2a_3a_4\}'. \end{aligned}$$

The application of this identity to the tensor $F_{i_1i_2i_3i_4}$ results in

$$(7)\qquad \begin{aligned} F_{i_1i_2i_3i_4} = {} & \frac{1}{4!}[\ A + B + C + D + E] \\ & + \frac{1}{64}[24A + 8B - 8D - 8E] \\ & + \frac{1}{144}[24A - 12C + 24E] \\ & + \frac{1}{64}[24A - 8B + 8D - 8E] \\ & + \frac{1}{4!}[\ A - B + C - D + E] \end{aligned}$$

where

$$
\begin{aligned}
A &= F_{i_1 i_2 i_3 i_4}, \\
B &= F_{i_2 i_1 i_3 i_4} + F_{i_3 i_2 i_1 i_4} + F_{i_4 i_2 i_3 i_1} + F_{i_1 i_3 i_2 i_4} + F_{i_1 i_4 i_3 i_2} + F_{i_1 i_2 i_4 i_3}, \\
C &= F_{i_2 i_3 i_1 i_4} + F_{i_3 i_1 i_2 i_4} + F_{i_2 i_4 i_3 i_1} + F_{i_4 i_1 i_3 i_2} + F_{i_3 i_2 i_4 i_1} + F_{i_4 i_2 i_1 i_3} \\
&\qquad + F_{i_1 i_3 i_4 i_2} + F_{i_1 i_4 i_2 i_3}, \\
D &= F_{i_2 i_3 i_4 i_1} + F_{i_3 i_4 i_2 i_1} + F_{i_2 i_4 i_1 i_3} + F_{i_4 i_1 i_2 i_3} + F_{i_3 i_1 i_4 i_2} + F_{i_4 i_3 i_1 i_2}, \\
E &= F_{i_2 i_1 i_4 i_3} + F_{i_3 i_4 i_1 i_2} + F_{i_4 i_3 i_2 i_1}.
\end{aligned}
$$

In the above decompositions we have assumed that for the tensor $F_{i_1 i_2 \ldots i_p}$ we have $p \leq n$, n being the dimension of the coordinate system with which we are working. Cases for which $p > n$ will be discussed in section 3.

Subsequently, we shall find it convenient to use an exponential notation when two or more adjacent α's are equal, and as heretofore, to omit zeros which may occur at the end of a partition. Thus $[2^2, 1^2]$ denotes the partition $[2, 2, 1, 1, 0, 0]$.

The following facts are well known. To every form of tableau $[\alpha]$ corresponds an irreducible representation of the symmetric group, which is usually denoted by the same symbol $[\alpha]$. The degree of this representation $[\alpha]$ is f_a, where f_a is given by Young's formula (2); or, alternatively stated, the f_a's are the numbers in the first column of the character tables referred to in section 2.

2. Immanant tensors. A tensor whose components are constant in all coordinate systems is called a numerical tensor.[18] With respect to an unrestricted transformation the simplest numerical tensor is the Kronecker delta

$$
\delta_j{}^i, \quad \text{where} \quad \delta_j{}^i \begin{cases} = 0, \text{ if } i = j, \\ = 1, \text{ if } i \neq j. \end{cases}
$$

A generalization of the simple Kronecker delta, which is called the generalized Kronecker delta, has m superscripts and m subscripts, each running from 1 to n, and is alternating in both superscripts and subscripts. It is denoted by

$$
\delta^{i_1 i_2 \cdots i_m}_{j_1 j_2 \cdots j_m} = \sum \delta_{j_1}{}^{i_1} \delta_{j_2}{}^{i_2} \cdots \delta_{j_m}{}^{i_m}, \qquad (m \leq n) \tag{8}
$$

where the $m!$ terms formed by permuting the i's are added, the sign being plus when the permutation is even, and minus when the permutation is odd. If the superscripts are distinct from each other and the subscripts are the same set of numbers as the superscripts, the value of the symbol is $+1$ or -1 according as an even or odd permutation is required to arrange the superscripts

[18] [23].

in the same order as the subscripts; in all other cases its value is zero. Veblen has given an interesting historical account of this symbol.[19] Murnaghan used it in the study of determinants,[20] and also proved its tensor character.[21] Cramlet carried the generalization still further, and found the most general absolute numerical tensor to be [22]

$$(9) \qquad \gamma^{i_1 i_2 \cdots i_p}_{j_1 j_2 \cdots j_p} = \sum_k A^{(k)} \delta_{j_1}{}^{i_1} \delta_{j_2}{}^{i_2} \cdots \delta_{j_p}{}^{i_p},$$

where the $p!$ terms formed by permuting the i's and multiplying each by an arbitrary scalar are added. The $p!$ scalars are distinguished by the $p!$ permutations of the i's. Cramlet does not explicitly state what range p may have in (9). However, his repeated remark that when the scalars are $+1$ or -1 according as the permutation is even or odd the general numerical tensor (9) becomes the generalized Kronecker delta

$$\delta^{i_1 i_2 \cdots i_m}_{j_1 j_2 \cdots j_m} \qquad (m \leq n)$$

considered by Murnaghan, and also used extensively by Cramlet [23] and Mitchell,[24] implies an unnecessary restriction on p. For an examination of Cramlet's argument reveals that it holds for any value of p, whether it be greater than, equal to, or less than n. To be sure, for a value of p greater than n the determinant tensor is a null tensor, as may be some other specializations of the general numerical tensor. But, as will be made clear later, some other specializations of the general numerical tensor are not null tensors for p greater than n.

When the scalars in (9) are all equal to $+1$ we have what Cramlet calls the permanent tensor, which is denoted by

$$\pi^{i_1 i_2 \cdots i_p}_{j_1 j_2 \cdots j_p}.$$

In order that our subsequent statements be clear, it seems desirable that we recall some ideas and terms commonly used in dealing with the symmetric group. Each permutation of the symmetric group on p letters may be written in an unique manner as a product of cycles, no letter appearing in more than one cycle. A permutation which factors into ρ_1 unary cycles, ρ_2 binary cycles, etc., is said to be of the class $(\rho) = (\rho_1, \rho_2, \cdots, \rho_p)$ of the permutation on p letters. The number of classes of the symmetric group is equal to the number of partitions of p into sums of positive integers, and is the same as the number of forms of Young tableaux $[\alpha]$. For, writing

---

[19] [23], p. 12.
[20] [24].
[21] [25].
[22] [20].
[23] [21], [22].
[24] [26].

$$\rho_1 + \rho_2 + \cdots \rho_p = \alpha_1;$$
$$\rho_2 + \cdots + \rho_p = \alpha_2; \quad \cdots; \quad \rho_p = \alpha_p;$$

it follows that

$$\alpha_1 + \alpha_2 + \cdots + \alpha_p = p;$$
$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_p.$$

It is customary to denote the class $(\rho) = (\rho_1, \rho_2, \cdots, \rho_n)$ by $(1^{\rho_1}, 2^{\rho_2}, 3^{\rho_3}, \cdots)$.

We now make the following

DEFINITION. *When the $p!$ scalars, the A's, of the general absolute numerical tensor*

(9)
$$\sum_k A^{(k)} \delta_{j_1}{}^{i_1} \delta_{j_2}{}^{i_2} \cdots \delta_{j_p}{}^{i_p}$$

are the characteristics of the symmetric group which corresponds to the partition $[\alpha]$ of p, the numerical tensor is said to be an immanant tensor, and is denoted by

(10)
$$_{\alpha}I^{i_1 i_2 \cdots i_p}_{j_1 j_2 \cdots j_p} = \sum_k \chi_{\alpha}{}^{(k)} \delta_{j_1}{}^{i_1} \delta_{j_2}{}^{i_2} \cdots \delta_{j_p}{}^{i_p}.$$

That is, for any partition $[\alpha]$ there is an immanant tensor, the $p!$ scalars in (9) for that partition being the characteristics associated with the class permutations for that partition. When no ambiguity arises, we may denote the immanant tensor by I_{α}. To facilitate the discussion of immanant tensors we list tables of characters [25] for several small values of p.

Tables of Characters of the Symmetric Groups.

Degree 2.

Class	(1^2)	(2)
Order	1	1
$[2]$	1	1
$[1^2]$	1	-1

Degree 3.

Class	(1^3)	$(1, 2)$	(3)
Order	1	3	2
$[3]$	1	1	1
$[2, 1]$	2	0	-1
$[1^3]$	1	-1	1

Degree 4.

Class	(1^4)	$(1^2, 2)$	$(1, 3)$	(4)	(2^2)
Order	1	6	8	6	3
$[4]$	1	1	1	1	1
$[3, 1]$	3	1	0	-1	-1
$[2^2]$	2	0	-1	0	2
$[2, 1^2]$	3	-1	0	1	-1
$[1^4]$	1	-1	1	-1	1

---

[25] [12], [18], [19].

For $p = 2 \leq n$, the only immanant tensors are the permanent tensor, $\pi^{i_1i_2}_{j_1j_2}$, and the determinant tensor, $\delta^{i_1i_2}_{j_1j_2}$. For interest in themselves and for clarity in subsequent remarks, we exhibit the components of these tensors for $n = 2$.

$\pi^{i_1i_2}_{j_1j_2}$	11	12	21	22
11	2	0	0	0
12	0	1	1	0
21	0	1	1	0
22	0	0	0	2

$\delta^{i_1i_2}_{j_1j_2}$	11	12	21	22
11	0	0	0	0
12	0	1	-1	0
21	0	-1	1	0
22	0	0	0	0

For $p = 3 \leq n$, in addition to the permanent tensor

$$I_{[3]} = \pi^{i_1i_2i_3}_{j_1j_2j_3},$$

and the determinant tensor

$$I_{[1^3]} = \delta^{i_1i_2i_3}_{j_1j_2j_3},$$

there is the immanant tensor

$$\begin{aligned} I_{[2,1]} &= \sum \chi^{(k)} \delta_{j_1}{}^{i_1} \delta_{j_2}{}^{i_2} \delta_{j_3}{}^{i_3} \\ &= 2\delta_{j_1}{}^{i_1}\delta_{j_2}{}^{i_2}\delta_{j_3}{}^{i_3} - \delta_{j_1}{}^{i_2}\delta_{j_2}{}^{i_3}\delta_{j_3}{}^{i_1} - \delta_{j_1}{}^{i_3}\delta_{j_2}{}^{i_1}\delta_{j_3}{}^{i_2}. \end{aligned}$$

When $n = 2$, $I_{[1^3]}$ is a null tensor, for we cannot have skew-symmetry on p things unless they are distinct. Put differently,

$$\delta^{i_1i_2i_3}_{j_1j_2j_3} = \begin{vmatrix} \delta_{j_1}{}^{i_1} & \delta_{j_1}{}^{i_2} & \delta_{j_1}{}^{i_3} \\ \delta_{j_2}{}^{i_1} & \delta_{j_2}{}^{i_2} & \delta_{j_2}{}^{i_3} \\ \delta_{j_3}{}^{i_1} & \delta_{j_3}{}^{i_2} & \delta_{j_3}{}^{i_3} \end{vmatrix},$$

and this determinant is obviously zero when $n = 2$, because two columns are equal.

For $p = 4 \leq n$, the five immanant tensors are:

$$\begin{aligned} I_{[4]} &= A + B + C + D + E; \\ I_{[3,1]} &= 3A + B - D - E; \\ I_{[2^2]} &= 2A - C - 2E; \\ I_{[2,1^2]} &= 3A - B + D - E; \\ I_{[1^4]} &= A - B + C - D + E; \end{aligned}$$

where

$$A = \delta_{j_1}{}^{i_1}\delta_{j_2}{}^{i_2}\delta_{j_3}{}^{i_3}\delta_{j_4}{}^{i_4} \quad \text{(the identity)},$$

$B =$ the sum of six terms for which the permutation of the i's is of the class $(1^2, 2)$,

$C =$ the sum of eight terms for which the permutation of the i's is of the class $(1, 3)$,

$D =$ the sum of six terms for which the permutation of the i's is of the class (4),

$E =$ the sum of three terms for which the permutation of the i's is of the class (2^2).

When $n = 3$, $I_{[1^4]}$ is a null tensor; and when $n = 2$, the last two of the five immanant tensors listed above are null tensors.

3. Decomposition of an arbitrary tensor with the use of immanant tensors. Since the decomposition of an arbitrary tensor has been defined by Weyl in terms of Young's symmetry operators, and since we have defined immanant tensors in terms of characters and permutations, the application of immanant tensors to the decomposition process depends on a relation between Young's symmetry operators and characters and permutations. Unfortunately, we are at a disadvantage in referring to the various parts of Young's analysis collectively, as we need to do here, because as Young developed his theory he found it desirable to make changes in method and in notation. Nevertheless, holding to the notation of section **1**, it may be said that the tableau $[\alpha]$ may be defined as standard if the letters in each row and column appear in the order of some given sequence. It can be shown that exactly f_α of the $p!$ tableaux are standard, where f_α is given by (2). Moreover, it can be shown that T_α may be expressed in terms of the standard tableaux, this expression being

$$T_\alpha = \frac{f_\alpha}{p!}\,[P_1N_1M_1 + P_2N_2M_2 + \cdots P_{f_\alpha}N_{f_\alpha}M_{f_\alpha}], \tag{11}$$

or

$$T_\alpha = \frac{f_\alpha}{p!}\sum_{r=1}^{f_\alpha} P_rN_rM_r,$$

where the multipliers M_r are introduced for orthogonality reasons.

Now we quote from Young [26] the

THEOREM I.

$$\frac{f_\alpha}{p!}\sum_{r=1}^{f_\alpha} P_rN_rM_r = \frac{f_\alpha}{p!}\,\Sigma\chi_\sigma\sigma \tag{12}$$

where χ_σ is the characteristic of the permutation of class σ.

Thus

$$T_\alpha = \frac{f_\alpha}{p!}\,\Sigma\,\chi_\sigma\sigma. \tag{13}$$

[26] [4], p. 256, Corollary to Theorem I.

From (13), and the definition of an immanant tensor, (10), we have

THEOREM II. *Operating on the i's we get*

(14) $$T_\alpha \cdot \delta_{j_1}{}^{i_1}\delta_{j_2}{}^{i_2} \cdots \delta_{j_p}{}^{i_p} = \frac{f_\alpha}{p!} I_\alpha.$$

Further, combining this with (4) we get

THEOREM III.

(15) $$p!\delta_{j_1}{}^{i_1}\delta_{j_2}{}^{i_2} \cdots \delta_{j_p}{}^{i_p} = \sum f_\alpha I_\alpha.$$

This is the basic identity satisfied by immanants.

With the tensor algebraic operation of contraction in mind it should be clear that if we multiply

$$F_{i_1 i_2 \ldots i_p}$$

by

$$_\alpha I^{i_1 i_2 \cdots i_p}_{j_1 j_2 \cdots j_p}$$

and contract the result will be an expression in the $F_{i_1 i_2 \ldots i_p}$'s for which the indices obey the same law as the i's in I_α. Thus we are led to

THEOREM IV. *A tensor $F_{i_1 i_2 \ldots i_p}$ is completely decomposed by multiplying the identity* (15) *through by it, and then contracting.*

Note that the elements of the tensor obtained by the process of section **1**, that is, of

$$\sum P_\alpha N_\alpha \cdot T_{i_1 i_2 \ldots i_p},$$

have the common multiplicative factor $\dfrac{p!}{f_\alpha}$. To illustrate, for $p = 4$, $f_{[3,1]} = 3$, and

$$\frac{p!}{f_\alpha} = \frac{24}{3} = 8;$$

which checks for the expression for $T_{[3,1]}F_{i_1 i_2 i_3 i_4}$ given by (7).

In the above applications we have considered the decomposition of an arbitrary covariant tensor $F_{i_1 i_2 \ldots i_p}$. However, Theorems II, III, and IV also constitute machinery for the decomposition of an arbitrary contravariant tensor $F^{j_1 j_2 \cdots j_p}$, if we think of the j's in (14) as being operated on, instead of the i's, as formerly done. Indeed, they form the basis for the decomposition of what Weyl has called " bisymmetric " tensors, tensors of the form

$$F^{i_1 i_2 \cdots i_p}_{j_1 j_2 \cdots j_p},$$

such that this tensor is left invariant if both the i's and j's are submitted to the same arbitrary permutation. So one can say that the immanant tensors are bisymmetry operators, as well as symmetry operators.

If we let

(16) $$I = \delta_{j_1}{}^{i_1}\delta_{j_2}{}^{i_2}\cdots\delta_{j_p}{}^{i_p}$$

and

(17) $$I'_\alpha = \frac{f_\alpha}{p!}\, I_\alpha,$$

the basic immanant tensor identity (15) becomes

(18) $$I = I'_{[p]} + \cdots + I'_{[1^p]},$$

or

$$I = \sum I'_\alpha :$$

Analogous to the properties of Young's T_α's

(19) $$T_\alpha \cdot T_\alpha = T_\alpha,$$

(20) $$T_\alpha \cdot T_\beta = 0 \qquad (\alpha \neq \beta),$$

we have from Theorem II the

THEOREM V.

(21) $$I'_\alpha \cdot I'_\alpha = I'_\alpha,$$

(22) $$I'_\alpha \cdot I'_\beta = a\ null\ tensor \qquad (\alpha \neq \beta),$$

where the multiplication is not what is conventionally called multiplication in tensor algebra, but is tensor multiplication followed by the contraction of a set of upper indices and a set of lower indices; it is convenient to refer to this as *bimultiplication* of the tensors. To illustrate, when $p = 2$,

(23) $$\pi^{i_1 i_2}_{l_1 l_2} \cdot \pi^{l_1 l_2}_{j_1 j_2} = 2\pi^{i_1 i_2}_{j_1 j_2} \; ; \; \delta^{i_1 i_2}_{l_1 l_2} \cdot \delta^{l_1 l_2}_{j_1 j_2} = 2\delta^{i_1 i_2}_{j_1 j_2},$$

(24) $$\pi^{i_1 i_2}_{l_1 l_2} \cdot \delta^{l_1 l_2}_{j_1 j_2} = \text{a null tensor.}$$

To exhibit in detail such a product as (24), let

$$A^{i_1 i_2}_{j_1 j_2} = \delta^{i_1 i_2}_{l_1 l_2} \cdot \pi^{l_1 l_2}_{j_1 j_2}.$$

Then from the tabular values of the factors previously given we construct the following table.

$A^{i_1 i_2}_{j_1 j_2}$	11	12	21	22
11	(0,0,0,0) .(2,0,0,0) = 0	(0,1,—1,0) .(2,0, 0,0) = 0	(0,—1,1,0) .(2, 0,0,0) = 0	(0,0,0,0) .(2,0,0,0) = 0
12	(0,0,0,0) .(0,1,1,0) = 0	(0,1,—1,0) .(0,1, 1,0) = 0	(0,—1,1,0) .(0, 1,1,0) = 0	(0,0,0,0) .(0,1,1,0) = 0
21	(0,0,0,0) .(0,1,1,0) = 0	(0,1,—1,0) .(0,1, 1,0) = 0	(0,—1,1,0) .(0, 1,1,0) = 0	(0,0,0,0) .(0,1,1,0) = 0
22	(0,0,0,0) .(0,0,0,2) = 0	(0,1,—1,0) .(0,0, 0,2) = 0	(0,—1,1,0) .(0, 0,0,2) = 0	(0,0,0,0) .(2,0,0,0) = 0

Just as the determinant tensor is a special case of an immanant tensor, the identity of classical tensor algebra [27]

$$\delta^{i_1 i_2 i_3}_{l_1 l_2 l_3} \cdot \delta^{l_1 l_2 l_3}_{j_1 j_2 j_3} = 3\,!\,\delta^{i_1 i_2 i_3}_{j_1 j_2 j_3}$$

is a special case of the immanant identity (21).

In virtue of the properties expressed in (21) and (22), the numerical tensors I_α's are "normal idempotents." With an arbitrary tensor one can generate a tensor having a prescribed type of symmetry through bimultiplication of this tensor with the appropriate I_α; the result is invariant under bimultiplication with I_α, and is annihilated through bimultiplication with I_β, where $\beta \neq \alpha$. Thus, since

$$\pi^{i_1 i_2}_{j_1 j_2} F_{i_1 i_2} = F_{j_1 j_2} + F_{j_2 j_1}, \qquad \delta^{i_1 i_2}_{j_1 j_2} F_{i_1 i_2} = F_{j_1 j_2} - F_{j_2 j_1},$$

$$\begin{aligned} \delta^{j_1 j_2}_{l_1 l_2}(\pi^{i_1 i_2}_{j_1 j_2} F_{i_1 i_2}) &= \delta^{j_1 j_2}_{l_1 l_2}(F_{j_1 j_2} + F_{j_2 j_1}), \\ &= F_{l_1 l_2} - F_{l_2 l_1} + F_{l_2 l_1} - F_{l_1 l_2}, \\ &= \text{a null tensor.} \end{aligned}$$

Similarly,

$$\begin{aligned} \pi^{j_1 j_2}_{l_1 l_2}(\pi^{i_1 i_2}_{j_1 j_2} F_{i_1 i_2}) &= \pi^{j_1 j_2}_{l_1 l_2}(F_{j_1 j_2} + F_{j_2 j_1}), \\ &= F_{l_1 l_2} + F_{l_2 l_1} + F_{l_2 l_1} + F_{l_1 l_2}, \\ &= 2(F_{l_1 l_2} + F_{l_2 l_1}). \end{aligned}$$

UNIVERSITY OF ALABAMA.

REFERENCES.

[1] H. Weyl, *The Classical Groups*, Princeton, 1939, pp. 96-136, especially section 4, Chapter IV.

[2] ———, *The Theory of Groups and Quantum Mechanics*, New York, 1932, Ch. V.

[3] A. Young, "On quantitative substitutional analysis," Part I, *Proceedings of the London Mathematical Society* (1), vol. 33 (1901), pp. 97-146; Part II, *ibid.* (1), vol. 34 (1902), pp. 361-397; Part III, *ibid.* (2), vol. 28 (1928); Also, "On quantitative substitutional analysis," *Journal London Mathematical Society*, vol. 3 (1928), pp. 14-19.

[4] ———, "On quantitative substitutional analysis," Part IV, *Proceedings of the London Mathematical Society* (2), vol. 31 (1930), pp. 253-272; especially Theorems I and II.

[5] ———, "On quantitative substitutional analysis," Part V, *ibid.* (2), vol. 31 (1930), pp. 273-288; Part VI, *ibid.* (2), vol. 34 (1932), pp. 196-230; Part VII, *ibid.* (2), vol. 36 (1933), pp. 304-368; Part VIII, *ibid.* (2), vol. 37 (1934), pp. 441-495.

[6] ———, "The application of substitutional analysis to invariants," *Philosophical Transactions of the Royal Society of London*, (A), vol. 234 (1935), pp. 79-114.

[27] [27].

[7] G. de B. Robinson, "On the geometry of the linear representation of the symmetric group," *Proceedings of the London Mathematical Society*, vol, 38 (1935), pp. 402-413.

[8] ———, "On the representations of the symmetric group," *American Journal of Mathematics*, vol. 60 (1938), pp. 745-760.

[9] G. Frobenius, "Über die Charaktere der symmetrischen Gruppe," *Berliner Berichte* (1903), pp. 328-358.

[10] I. Schur, "Neue Begründung der Theorie der Gruppentheorie," *Berliner Berichte*, (1905), pp. 406-432.

[11] W. Burnside, *Theory of Groups of Finite Order*, Cambridge, 1911.

[12] D. E. Littlewood and A. R. Richardson, "Group characters and algebra," *Philosophical Transactions of the Royal Society of London*, (A), vol. 233 (1934), pp. 99-141.

[13] ———, "Immanants of some special matrices," *Quarterly Journal of Mathematics*, vol. 5 (1934), pp. 269-282.

[14] D. E. Littlewood, "Group characters and the structure of groups," *Proceedings of the London Mathematical Society* (2), vol. 39 (1935), pp. 150-199.

[15] M. Zia-ud-Din, "The characters of the symmetric group of order 11," *ibid.* (2) vol. 39 (1935), pp. 200-204.

[16] ———, "The characters of the symmetric groups of degrees 12 and 13," *ibid.* (2), vol. 42 (1937), pp. 340-355.

[17] F. D. Murnaghan, "On the representations of the symmetric group," *American Journal of Mathematics*, vol. 59 (1937), pp. 437-488.

[18] ———, "The characters of the symmetric group," *ibid.*, vol. 59 (1937), pp. 739-753.

[19] ———, *The Theory of Group Representations*, Baltimore, 1938, Ch. III.

[20] C. M. Cramlet, "A determination of all invariant tensors," *Tohoku Mathematical Journal*, vol. 28 (1927), pp. 242-250.

[21] ———, "The derivation of algebraic invariants by tensor algebra," *Bulletin of the American Mathematical Society*, vol. 34 (1928), pp. 334-342.

[22] ———, "The application of the determinant and permanent tensors to determinants of general class and allied tensor functions," *American Journal of Mathematics*, vol. 49 (1927), pp. 87-96.

[23] O. Veblen, *Invariants of Quadratic Differential Forms, Cambridge Tracts* no. 24, 1927, p. 25.

[24] F. D. Murnaghan, "The generalized Kronecker symbol and its application to the theory of determinants," *American Mathematical Monthly*, vol. 32 (1925), pp. 233-241.

[25] ———, "The tensor character of the generalized Kronecker symbol," *Bulletin of the American Mathematical Society*, vol. 31 (1925), pp. 323-334.

[26] A. K. Mitchell, "The derivation of tensors from tensor functions," *American Journal of Mathematics*, vol. 49 (1929), pp. 87-96.

[27] A. J. McConnell, *Applications of the Absolute Differential Calculus*, Blackie and son, 1931, p. 17.

EQUIVALENCE OF QUADRATIC FORMS.*

By Carl Ludwig Siegel.

Introduction. Let S be a quadratic form in m variables $x_1, \cdots, x_m$ with coefficients belonging to a ring P and T a quadratic form in n variables $y_1, \cdots, y_n$. We say that S *represents* T *in* P if S is carried into T by a linear transformation

$$x_k = \sum_{l=1}^{n} c_{kl} y_l \qquad (k = 1, \cdots, m), \tag{1}$$

where the coefficients c_{kl} are numbers of P. If also T represents S in P, we say that S and T are *equivalent in* P. We shall assume the ring P to be one of the rings R, R_p, R_∞, J, J_p defined in the following manner: R is the field of rational numbers, R_p the field of p-adic numbers, where p denotes any prime number, R_∞ the field of real numbers; moreover J is the ring of integral numbers and J_p the ring of p-adic integers. It is convenient to denote R_∞ also by J_∞. If we speak of *equivalence* without mentioning the ring, we always mean equivalence in J. On the other hand, equivalence in R is also called *rational equivalence.*

Since R is contained in all the fields R_p and R_∞, two rationally equivalent quadratic forms are certainly also equivalent in all R_p and R_∞. The converse of this obvious statement is

Theorem 1. *If two quadratic forms with rational coefficients are equivalent in all R_p and R_∞, then they are also rationally equivalent.*

This important theorem was found by Minkowski[1] who published only a sketch of the proof. The first detailed proof was given by Hasse.[2]

We consider now the corresponding questions for J, J_p, J_∞ instead of R, R_p, R_∞. It is again obvious that two equivalent quadratic forms are *a fortiori* equivalent in all J_p and J_∞. However it may be seen from simple examples that now the converse is not generally true. In order to overcome this difficulty we introduce the notion of semiequivalence:

* Received March 5, 1941.

[1] H. Minkowski, *Gesammelte Abhandlungen*, vol. 1, p. 222.

[2] H. Hasse, "Über die Äquivalenz quadratischer Formen im Körper der rationalen Zahlen," *Journal für die reine und angewandte Mathematik*, vol. 152 (1923), pp. 205-224.

Let S and T be quadratic forms with integral coefficients. We say that *S represents T rationally without essential denominator,* if there exists for any positive integer q a linear transformation (1) carrying S into T, such that the coefficients c_{kl} $(k=1,\cdots,m;\ l=1,\cdots,n)$ are rational numbers and their denominators relative-prime to q. If also T represents S rationally without essential denominator, then S and T are called *semiequivalent.* Obviously the equivalence in all J_p and J_∞ follows already from semiequivalence. And now also the converse holds, namely

THEOREM 2. *If two quadratic forms with integral coefficients are equivalent in all J_p and J_∞, then they are also semiequivalent.*

The *genus* of a quadratic form S with integral coefficients is the set of the quadratic forms T equivalent to S in all J_p and J_∞. Theorem 2 maintains that two quadratic forms of the same genus always represent each other rationally without essential denominator. This has been proved by Smith [3] in the case of three variables. The general theorem was stated by Minkowski [4]; but a complete proof had hitherto not been given.

The assumptions in Theorems 1 and 2 contain only local properties of the quadratic forms dealing with their behavior at the finite and infinite prime-spots; on the other hand, the assertion is the existence of a certain matrix in the large, namely in the field of rational numbers. It is therefore quite natural that the proofs depend upon the transcendental methods of analytic number-theory. Hitherto the proofs of Theorem 1 used Dirichlet's theorem concerning the prime numbers in arithmetical progressions. This theorem, however, belongs to the theory of class-fields rather than to the proper theory of quadratic forms. It is of methodical interest to refrain from using Dirichlet's theorem and to apply in its place the properties of the theta functions giving the adequate tools for the investigation of quadratic forms.

Our proofs of the two theorems require very little knowledge of arithmetics. We use only the properties of linear forms contained in the main theorem on elementary divisors of rational matrices. We do not need the law of quadratic reciprocity which was important in the former proofs of Theorem 1. The method of our proofs could be generalized without difficulty to the case of quadratic forms in arbitrary algebraic number-fields.

In the statement of Theorems 1 and 2 we did not assume that the two quadratic forms are non-degenerate, i. e. that their determinants are different

[3] H. J. St. Smith, *Collected Papers*, vol. 1, p. 480.

[4] H. Minkowski, *Gesammelte Abhandlungen*, vol. 1, p. 221.

from 0. We show now by a simple consideration that we may restrict ourselves to the non-degenerate case.

The rank of the product of two matrices is not greater than the rank of either of the factors. Hence two quadratic forms S and T have obviously the same rank μ, if they are equivalent in P. Since Theorems 1 and 2 are certainly true in the case $\mu = 0$, we may assume $\mu > 0$. We want to prove that S is equivalent in P to a non-degenerate quadratic form in μ variables.

Let us first consider the cases $\mathrm{P} = R, R_p, R_\infty$. If the m coefficients s_{kk} $(k = 1, \cdots, m)$ of

$$S = \sum_{k,l=1}^{m} s_{kl} x_k x_l$$

are all 0, then at least one of the other coefficients $s_{kl} = s_{lk}$ $(k \neq l)$ is $\neq 0$, say s_{ab}. The special linear transformation $x_b \to x_a + x_b$, $x_k \to x_k$ $(k \neq b)$ and its inverse belong to P; hence S is equivalent in P to $S + 2x_a \sum_{k=1}^{m} s_{kb} x_k$, and the coefficient of x_a^2 in this quadratic form is $2s_{ab} \neq 0$. Hence we may assume that already one of the coefficients s_{kk}, say s_{aa}, is $\neq 0$. Then the difference

$$S - s_{aa}(x_a + \sum_{k \neq a} \frac{s_{ak}}{s_{aa}} x_k)^2$$

is a quadratic form in the $m - 1$ variables x_k $(k \neq a)$. Applying induction we infer that S is equivalent in P to a *diagonal form* $c_1 x_1^2 + \cdots + c_\mu x_\mu^2$, where the coefficients $c_1, \cdots, c_\mu$ are $\neq 0$. In the case $\mathrm{P} = R$ or R_p, let c be a common denominator of the numbers $c_1, \cdots, c_\mu$. By the substitutions $x_k \to c x_k$ $(k = 1, \cdots, \mu)$, we get integral coefficients. Consequently we may assume for the proof of Theorem 1, that the two quadratic forms are diagonal forms with integral coefficients.

In the cases $\mathrm{P} = J, J_p$, we apply the theory of elementary divisors to the matrix $\mathfrak{S} = (s_{kl})$ of the quadratic form S. There exist two matrices $\mathfrak{U}$ and $\mathfrak{V}$ of P with m rows and determinant 1, such that

$$\mathfrak{U}\mathfrak{S}\mathfrak{V} = \begin{pmatrix} \mathfrak{D} & 0 \\ 0 & 0 \end{pmatrix},$$

where $\mathfrak{D}$ is a diagonal matrix of μ rows and 0 is a zero matrix. If $\mathfrak{U}'$ denotes the transpose of $\mathfrak{U}$, the matrix

$$\mathfrak{U}\mathfrak{S}\mathfrak{U}' = \begin{pmatrix} \mathfrak{D} & 0 \\ 0 & 0 \end{pmatrix} \mathfrak{V}^{-1}\mathfrak{U}'$$

is symmetric. Since the elements of the last $m - \mu$ rows on the right-hand side are all 0, we obtain

$$\mathfrak{U}\mathfrak{S}\mathfrak{U}' = \begin{pmatrix} \mathfrak{W} & 0 \\ 0 & 0 \end{pmatrix},$$

where $\mathfrak{W}$ is a symmetric matrix of μ rows with non-vanishing determinant. Hence S is equivalent in $\mathbf{P}$ to the quadratic form with the matrix $\mathfrak{W}$, and we may assume for the proof of Theorem 2 that the two quadratic forms are non-degenerate.

Henceforth we denote by S only non-degenerate quadratic forms. For our further researches the notion of the zero-form is important: A quadratic form S with coefficients in $\mathbf{P}$ is called a *zero-form in* $\mathbf{P}$, if the Diophantine equation $S=0$ has a solution in numbers $x_1, \cdots, x_m$ of $\mathbf{P}$ which are not all 0. A zero-form in R is obviously also a zero-form in all R_p and R_∞. Again the converse is true:

A quadratic form with rational coefficients is a zero-form in R, if it is a zero-form in all R_p and R_∞.

This was proved for the case $m=3$ by Legendre,[5] in a different notation, and by Hasse [6] for the case of any m. If $m>4$, the theorem is contained in the following result of A. Meyer [7]:

Any indefinite quadratic form with rational coefficients and more than 4 *variables is a zero-form in R.*

Since a quadratic form S with rational coefficients is equivalent in R to a diagonal form $a_1x_1{}^2 + \cdots + a_mx_m{}^2$ with integral coefficients $a_k \neq 0$ $(k=1, \cdots, m)$, we may assume for the proof of these theorems of Hasse-Legendre and Meyer that S is such a diagonal form. On the other hand, the equation $S=0$ is homogeneous and we may restrict the variables $x_1, \cdots, x_m$ to integral values. In addition to the indefinite form

$$S = \sum_{k=1}^{m} a_k x_k{}^2$$

we introduce the positive quadratic form

$$\mathfrak{P} = \sum_{k=1}^{m} |\, a_k \,|\, x_k{}^2$$

and consider for arbitrary $\epsilon > 0$ the sum

$$A(\epsilon) = \sum_{S=0} \exp(-\pi\epsilon\mathfrak{P}) \tag{2}$$

extended over all integral solutions $x_1, \cdots, x_m$ of $S=0$.

---

[5] A. M. Legendre, *Théorie des nombres*, ed. 3 (1830), vol. 1, §§ 3, 4.

[6] H. Hasse, "Über die Darstellbarkeit von Zahlen durch quadratische Formen im Körper der rationalen Zahlen," *Journal für die reine und angewandte Mathematik*, vol. 152 (1923), pp. 129-148.

[7] A. Meyer, "Über die Kriterien für die Auflösbarkeit der Gleichung $ax^2 + by^2 + cz^2 + du^2 = 0$ in ganzen Zahlen," *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, vol. 29 (1884), p. 209-222.

THEOREM 3. *Let S be an indefinite quadratic diagonal form, with integral coefficients, in more than 4 variables. If ϵ tends to 0 through positive values, then the expression $\epsilon^{(m/2)-1}A(\epsilon)$ tends to a positive limit.*

This is obviously a quantitative improvement of Meyer's theorem. We shall give a proof using the "circle method" of Hardy and Littlewood. The same method leads to the following special result in the case $m = 4$:

THEOREM 4. *Let S be an indefinite quaternary quadratic diagonal form with integral coefficients. If S is a zero-form in all R_p and if its determinant is the square of an integer, then the expression $(\epsilon/\log \epsilon^{-1})A(\epsilon)$ tends for $\epsilon \to 0$ to a positive limit.*

This is a quantitative improvement of the Hasse-Legendre theorem, in the case of a quaternary form with quadratic determinant. It is also possible to determine the asymptotic behavior of $A(\epsilon)$ in the remaining cases, namely $m < 4$, or $m = 4$ and the determinant not a square. But in these cases the estimation of certain quantities in the proof becomes more laborious and requires particular calculations which are not necessary under the assumptions of Theorems 3 or 4. For the purpose of our investigation, the proof of Theorems 1 and 2, we need only the statements of Theorems 3 and 4; therefore we omit here the more difficult discussion of those particular cases.

1. Gaussian sums and singular series.

LEMMA 1. *Let ρ be a primitive root of unity of degree $q \geq 1$, α and β be two integers and $\delta = (2\alpha, q)$ be the greatest common divisor of 2α and q. Then*

$$\Big| \sum_{h=1}^{q} \rho^{\alpha h^2+\beta h} \Big|^2 \leq \delta q,$$

where the sign of equality is true in the case $\delta = 1$.

Proof. Multiplying the two conjugate complex quantities

$$G = \sum_{h=1}^{q} \rho^{\alpha h^2+\beta h}, \qquad \bar{G} = \sum_{k=1}^{q} \rho^{-\alpha k^2-\beta k}$$

we find

$$| G |^2 = \sum_{k=1}^{q} \rho^{-\alpha k^2-\beta k} \Big(\sum_{h=k+1}^{q+k} \rho^{\alpha h^2+\beta h} \Big) = \sum_{l=1}^{q} \rho^{\alpha l^2+\beta l} \Big(\sum_{k=1}^{q} \rho^{2\alpha k l} \Big).$$

The inner sum on the right-hand side is 0, if $2\alpha l$ is not a multiple of q, and q otherwise. Since q is a factor of $2\alpha l$ only for the δ values $l = gq\delta^{-1}$ ($g = 1, \cdots, \delta$) in the interval $1 \leq l \leq q$, we have

$$| G |^2 = q \sum_{l}{}' \rho^{\alpha l^2+\beta l},$$

the sum being extended over these values of l, and hence

$$| G |^2 \leq \delta q,$$

where the sign of equality certainly holds in the case $\delta = 1$.

In this paragraph we denote by S a diagonal form

$$S = S(x) = \sum_{k=1}^{m} a_k x_k^2$$

with integral coefficients a_k, where x is the row of the variables $x_1, \cdots, x_m$. Let

$$D = a_1 \cdots a_m$$

be the determinant of S. For any positive integer q, a special primitive q-th root of unity is given by

$$\rho_q = e^{2\pi i/q}.$$

The Gaussian sums are defined by

$$G_q(r) = \sum_{x(q)} \rho_q^{rS(x)}, \tag{3}$$

where the sign $x(q)$ indicates that the variables $x_1, \cdots, x_m$ run independently over a complete system of residues modulo q, and r is an integral number. If t is a common divisor of $q = tq_1$ and $r = tr_1$, we have obviously

$$G_q(r) = t^m G_{q_1}(r_1). \tag{4}$$

On the other hand, we infer from Lemma 1 the inequality

$$| G_q(r) |^2 \leq q^m \prod_{k=1}^{m} (2a_k r, q). \tag{5}$$

We define

$$H_q = q^{-m} \sum_{r(q)}' G_q(r), \tag{6}$$

where the sign $r(q)$ and the dash mean that r runs through a reduced system of residues modulo q. Denoting by $\phi(q)$ Euler's function, we obtain by (5) the estimate

$$| H_q | \leq \phi(q) q^{-m/2} \prod_{k=1}^{m} (2a_k, q)^{\frac{1}{2}}, \tag{7}$$

whence

$$| H_q | \leq \phi(q) q^{-m/2}, \tag{8}$$

if $(2D, q) = 1$, and in any case

$$| H_q | < 2^{m/2} | D |^{\frac{1}{2}} q^{1-(m/2)}. \tag{9}$$

We assume now $m > 2$ and denote by p any prime number. By (9), the series

$$\sigma_p = \sum_{k=0}^{\infty} H_{p^k} \tag{10}$$

is absolutely convergent. On the other hand, let N_q be the number of solutions of the congruence

$$S(x) \equiv 0 \pmod q$$

in different systems $x(q)$. The significance of the quantity σ_p is made clear by

LEMMA 2. *If $q = p^l$ runs over the powers of the prime number p, then*

$$\lim_{q\to\infty} q^{1-m} N_q = \sigma_p. \tag{11}$$

This limit σ_p is $\neq 0$, if and only if S is a zero-form in R_p.

Proof. Let r_k run over a reduced system of residues modulo p^k $(k = 0, \cdots, l)$. The numbers $r_k p^{l-k}$ $(k = 0, \cdots, l)$ form exactly a complete system of residues for the modulus $p^l = q$. By (4), the relationship

$$G_{p^l}(r_k p^{l-k}) = p^{(l-k)m} G_{p^k}(r_k)$$

holds, and hence by (6)

$$\sum_{h=1}^{q} G_q(h) = \sum_{k=0}^{l} p^{(l-k)m} \sum_{r_k(p^k)}{}' G_{p^k}(r_k) = q^m \sum_{k=0}^{l} H_{p^k}.$$

On the other hand, by (3)

$$\sum_{h=1}^{q} G_q(h) = \sum_{x(q)} \sum_{h(q)} \rho_q^{hS(x)} = qN_q.$$

This leads to

$$q^{1-m} N_q = \sum_{k=0}^{l} H_{p^k} \tag{12}$$

and consequently to the first assertion of our lemma.

Now we shall prove that $\sigma_p > 0$, if S is a zero-form in R_p. Let p^α be the largest power of p dividing $2D$ and $l \geq 2\alpha + 1$, $q = p^l$. Since S is a zero-form in R_p, the congruence

$$S(x) \equiv 0 \pmod q \tag{13}$$

has a *primitive* solution x, i. e. a solution in integers $x_1, \cdots, x_m$ which are not all divisible by p. The greatest common divisor

$$(q, 2a_1x_1, \cdots, 2a_mx_m) = p^\beta \tag{14}$$

is then a factor of the number $(2D, q) = p^\alpha$, hence $\beta \leq \alpha$ and $2(l - \beta) \geq 2l - 2\alpha \geq l + 1$. If y is integral, we find

$$S(x + p^{l-\beta}y) \equiv S(x) + 2p^{l-\beta} \sum_{g=1}^{m} a_g x_g y_g \pmod{pq};$$

hence the number of modulo pq incongruent solutions z of

$$S(z) \equiv 0 \pmod{pq}, \qquad z \equiv x \pmod{p^{l-\beta}}$$

is the same as the number of modulo $p^{\beta+1}$ incongruent solutions y of

$$2p^{-\beta} \sum_{g=1}^{m} a_g x_g y_g \equiv -p^{-l} S(x) \pmod{p}.$$

By (13) and (14), this number is $p^{m-1+m\beta}$. On the other hand, the number of modulo q incongruent solutions z of

$$S(z) \equiv 0 \pmod{q}, \qquad z \equiv x \pmod{p^{l-\beta}}$$

is obviously $p^{m\beta}$, and the ratio of the two numbers of solutions is p^{m-1}, independent of the primitive solution x. Denoting by M_{p^k} the number of modulo p^k incongruent primitive solutions x of $S(x) \equiv 0 \pmod{p^k}$, we obtain by summation over x modulo $p^{l-\beta}$ the formula

$$M_{pq} = p^{m-1} M_q \qquad (q = p^l;\ l \geq 2\alpha + 1).$$

This shows that the expression $p^{(1-m)l} M_{p^l}$ $(l \geq 2\alpha + 1)$ is a positive number not depending upon l. But

$$N_q \geq M_q \qquad (q = p^l;\ l = 0, 1, \cdots),$$

and hence the sequence $q^{1-m} N_q$ has a positive lower bound. By (11), the inequality $\sigma_p > 0$ follows.

To complete the proof of our lemma, let us now assume, on the other hand, that S is not a zero-form in R_p. Then we infer, from the above discussion, that the congruence

$$S(x) \equiv 0 \pmod{q} \tag{15}$$

has no primitive solution, if $q = p^l$, $l \geq 2\alpha + 1$ and p^α is the largest power of p dividing $2D$. Let x be any integral solution of (15) and

$$(q, x_1, \cdots, x_m) = p^\gamma. \tag{16}$$

Then $y = p^{-\gamma} x$ satisfies the congruence

$$p^{2\gamma} S(y) \equiv 0 \pmod{q}.$$

Hence $2\gamma \geq l - 2\alpha$, and by (16)

$$N_q \leq p^{m(l-l/2+\alpha)} = p^{m\alpha} q^{m/2}$$
$$q^{1-m} N_q \leq p^{m\alpha} q^{1-(m/2)}.$$

Since $m > 2$, the right-hand side tends to zero for $q = p^l \to \infty$, and we see by (11) that $\sigma_p = 0$.

LEMMA 3. *Let either $m > 4$ and p arbitrary or $m > 2$ and $(p, 2D) = 1$; then*

$$\sigma_p > 0. \tag{17}$$

In the case $m > 4$, $(p, 2D) = 1$ the stronger inequality

$$\sigma_p > 1 - p^{-3/2} \tag{18}$$

holds.

Proof. We consider first the case $m > 2$, $(p, 2D) = 1$. For $q = p^l$ $(l = 1, 2, \cdots)$, the inequality (8) implies

$$| H_q | \leq (1 - p^{-1}) q^{1-m/2}$$

$$| \sum_{l=1}^{\infty} H_{p^l} | \leq (1 - p^{-1}) \sum_{l=1}^{\infty} p^{l(1-m/2)} = p^{1-m/2} \frac{1 - p^{-1}}{1 - p^{1-m/2}}.$$

The right-hand side of the last formula is $< p^{-3/2}$ for $m \geq 5$, $= p^{-1}$ for $m = 4$, $= p^{-1/2} + p^{-1}$ for $m = 3$, and less than 1 in any of these cases, since $p \geq 3$. Now $H_1 = 1$ and therefore

$$| \sigma_p | \geq 1 - | \sum_{l=1}^{\infty} H_{p^l} |.$$

By (11), the number σ_p is non-negative. Hence (17) holds for $m > 2$, $(p, 2D) = 1$, and (18) for $m > 4$, $(p, 2D) = 1$.

It remains to prove (17) for $m > 4$ and the prime factors p of $2D$. Applying Lemma 2 we have only to show that S is a zero-form in R_p. For this proof we may assume, without loss of generality, that

$$S = \sum_{k=1}^{h} a_k x_k^2 + p \sum_{k=h+1}^{m} b_k x_k^2, \tag{19}$$

where h is a certain number of the interval $0 \leq h \leq m$ and none of the integral coefficients a_k, b_k is divisible by p. Moreover we may restrict ourselves to the case $h \geq m/2$, since the case $h \leq m/2$ is transformed into this one, if we divide S by p and replace x_k by px_k $(k = 1, \cdots, h)$. If $p \neq 2$, we apply Lemma 2 and the already proved part of our lemma to the quadratic form

$$S_1 = \sum_{k=1}^{h} a_k x_k^2$$

in more than 2 variables. Then S_1 is a zero-form in R_p and consequently so also is S. In the remaining case $p = 2$ we infer from (7) and (19) the estimate

$$| H_q | \leq 2^{m-h/2-1} q^{1-m/2} \qquad (q = 2^l;\ l = 2, 3, \cdots).$$

Since $h > 0$, we find by direct calculation that $H_2 = 0$, and in the case $h < m$ also $H_4 = 0$. Hence

$$| \sum_{l=1}^{\infty} H_{2^l} | \leq 2^{m-h/2-1} \sum_{l=\lambda}^{\infty} 2^{l(1-m/2)} < 2^{m-h/2+\lambda(1-m/2)},$$

where $\lambda = 3$ for $h < m$ and $\lambda = 2$ for $h = m$. Now $h \geq m/2$ and $m > 4$, whence $m - h/2 + \lambda(1 - m/2) < 0$ in both cases and

$$| \sigma_2 | \geq 1 - | \sum_{l=1}^{\infty} H_{2^l} | > 0.$$

By Lemma 2, S is a zero-form in R_2.

LEMMA 4. *If $m > 4$, the series*

$$\sigma = \sum_{q=1}^{\infty} H_q \tag{20}$$

converges and

$$\sigma > 0. \tag{21}$$

Proof. Let $q = q_1 q_2$ be a decomposition of q into relative-prime factors. If $x^{(1)}$ and $x^{(2)}$ run over complete systems of residues mod q_1 and mod q_2, then $x = q_2 x^{(1)} + q_1 x^{(2)}$ runs over a complete system of residues mod q. In an analogous manner we may take $r = q_2 r_1 + q_1 r_2$, where r_1 and r_2 run over reduced systems of residues mod q_1 and mod q_2. We obtain

$$rS(x) = (q_2 r_1 + q_1 r_2) S(q_2 x^{(1)} + q_1 x^{(2)}) \equiv r_1 q_2^3 S(x^{(1)}) + r_2 q_1^3 S(x^{(2)}) \pmod q$$

$$\rho_q^{rS(x)} = \rho_{q_1}^{r_1 q_2^2 S(x^{(1)})} \rho_{q_2}^{r_2 q_1^2 S(x^{(2)})}$$

and by (3)

$$G_q(r) = G_{q_1}(r_1 q_2^2) G_{q_2}(r_2 q_1^2).$$

Since also $r_1 q_2^2$ and $r_2 q_1^2$ run over reduced systems mod q_1 and mod q_2, we find by (6)

$$H_q = H_{q_1} H_{q_2}. \tag{22}$$

By (9), the series (20) is absolutely convergent for $m > 4$. Hence from (10) and (22) the product-formula

$$\sigma = \prod_p \sigma_p$$

follows. By Lemma 3, all factors σ_p are positive, and moreover the inequality

$$\prod_{(p,2D)=1} \sigma_p > \prod_{(p,2D)=1} (1 - p^{-3/2}) > 0$$

holds, where the multiplication extends over all prime numbers p not dividing $2D$. Consequently (21) is proved.

LEMMA 5. *Let p be an odd prime number not dividing the determinant D; then S is equivalent in J_p to $Dx_m^2 + \sum_{k=1}^{m-1} x_k^2$.*

Proof. Let a and b be two integers not divisible by p. By (8) and (12), the number of modulo p incongruent solutions ξ, η, ζ of

$$a\xi^2 + b\eta^2 \equiv \zeta^2 \pmod{p} \tag{23}$$

is not less than $p^2\{1 - (1 - p^{-1})p^{-1/2}\}$, and the number of modulo p incongruent solutions ξ, η of $a\xi^2 + b\eta^2 \equiv 0 \pmod{p}$ is not more than $p(1 + 1 - p^{-1})$. Since the difference of both numbers is $\geq (p-1)(p - p^{1/2} - 1) > 0$, there exists a solution of (23) with $\zeta \not\equiv 0 \pmod{p}$. The method of the proof of Lemma 2 leads then to an integral p-adic solution α, β of $a\alpha^2 + b\beta^2 = 1$, and we get the identity

$$ax_1^2 + bx_2^2 = y_1^2 + aby_2^2,$$

where x_1, x_2 and y_1, y_2 are connected by the linear substitution

$$x_1 = \alpha y_1 - b\beta y_2, \qquad x_2 = \beta y_1 + a\alpha y_2$$

of determinant 1. Applying this result $m - 1$ times, we obtain the proof of the assertion.

For the rest of this paragraph, we assume that $m = 4$ and the determinant D of S is the square of an integer, $D = a^2$. Denoting by s a real variable, we define

$$\sigma_p(s) = \sum_{l=0}^{\infty} H_{p^l} p^{-ls}, \tag{24}$$

$$\sigma(s) = \sum_{q=1}^{\infty} H_q q^{-s}, \tag{25}$$

where the first sum, by (9), converges absolutely for $s > -1$ and the second sum for $s > 0$. By (22)

$$\sigma(s) = \prod_p \sigma_p(s) \qquad (s > 0), \tag{26}$$

and by (10)

$$\lim_{s \to 0} \sigma_p(s) = \sigma_p(0) = \sigma_p. \tag{27}$$

LEMMA 6. *If $m = 4$, $D = a^2$, $(p, 2D) = 1$, then*

$$\sigma_p(s) = \frac{1 - p^{-s-2}}{1 - p^{-s-1}} \qquad (s > -1).$$

Proof. By Lemma 5, the quadratic form S is equivalent in J_p to $x_1^2 + x_2^2 + x_3^2 + a^2x_4^2$, hence also to $x_1^2 + x_2^2 + x_3^2 + x_4^2$, and the same holds for the quadratic form

$$T = x_1^2 + x_2^2 - x_3^2 - x_4^2.$$

If two rows of variables are connected by a linear substitution with integral coefficients whose determinant is relative-prime to q, then they run at the same time over a complete system of residues modulo q. Hence, for $q = p^l$, the Gaussian sum $G_q(r)$ is not changed, if S is replaced by T. Applying the definition (3) we find

$$G_q(r) = (\sum_{h=1}^{q} \rho_q^{rh^2})^2 (\sum_{h=1}^{q} \rho_q^{-rh^2})^2 \qquad (q = p^l;\ l = 0, 1, \cdots)$$

and therefore by Lemma 1

$$G_q(r) = q^2,$$

if $(p, r) = 1$. By (6) and (24)

$$H_q = (1 - p^{-1}) q^{-1} \qquad (q = p^l;\ l = 1, 2, \cdots)$$

$$\sigma_p(s) = 1 + (1 - p^{-1}) \sum_{l=1}^{\infty} p^{-l(s+1)} = 1 + \frac{1 - p^{-1}}{p^{s+1} - 1} = \frac{1 - p^{-s-2}}{1 - p^{-s-1}}.$$

LEMMA 7. *If $m = 4$, $D = a^2$ and S is a zero-form in all R_p, then the sequence $\sum_{q=1}^{t} H_q / \sum_{q=1}^{t} q^{-1}$ $(t = 1, 2, \cdots)$ tends for $t \to \infty$ to a positive limit.*

Proof. By (26) and Lemma 6

$$\sigma(s) = \frac{\zeta(s+1)}{\zeta(s+2)} \prod_{p/2D} \frac{(1 - p^{-s-1}) \sigma_p(s)}{1 - p^{-s-2}} \qquad (s > 0).$$

The Dirichlet series

$$\frac{1}{\zeta(s+2)} \prod_{p/2D} \frac{(1 - p^{-s-1}) \sigma_p(s)}{1 - p^{-s-2}} = \psi(s) = \sum_{q=1}^{\infty} c_q q^{-s}$$

is the product of a finite number of Dirichlet series which are absolutely convergent for $s > -1$; hence it is also absolutely convergent for $s > -1$. In particular, for $s = 0$,

(28) $$\frac{6}{\pi^2} \prod_{p/2D} \frac{\sigma_p(0)}{1 + p^{-1}} = \psi(0) = \sum_{q=1}^{\infty} c_q.$$

Defining

$$\gamma(u) = \sum_{q > u} c_q$$

we have

$$\gamma(u) \to 0 \qquad (u \to \infty).$$

If δ is an arbitrarily small positive number, the inequality $|\gamma(u)| < \delta$ holds for $u > \nu = \nu(\delta) > 0$; moreover a constant K exists, such that $|\gamma(u)| < K$ for all values of u. Using the abbreviation

$$\sum_{q=1}^{t} q^{-1} = L_t \qquad (t = 1, 2, \cdots)$$

we find

$$\Big|\sum_{q=1}^{t}\gamma(t/q)q^{-1}\Big| \leq K \sum_{t/\nu\leq q\leq t} q^{-1} + \delta\sum_{q\leq t} q^{-1} \leq K\nu + \delta L_t$$

$$L_t^{-1}\sum_{q=1}^{t}\gamma(t/q)q^{-1} \to 0 \qquad (t\to\infty). \tag{29}$$

From (25) and the identity

$$\sigma(s) = \zeta(s+1)\psi(s) = \sum_{q=1}^{\infty} q^{-s-1}\sum_{q=1}^{\infty} c_q q^{-s}$$

we obtain

$$\sum_{q=1}^{t} H_q = \sum_{ab\leq t} c_a b^{-1} = \sum_{q=1}^{t}\{\psi(0)-\gamma(t/q)\}q^{-1} = \psi(0)L_t - \sum_{q=1}^{t}\gamma(t/q)q^{-1};$$

hence by (29)

$$L_t^{-1}\sum_{q=1}^{t} H_q \to \psi(0) \qquad (t\to\infty).$$

By (27), (28) and Lemma 2, the inequality $\psi(0) > 0$ holds.

2. Farey dissection and theta functions. Let ϵ be a given number,

$$0<\epsilon<1, \qquad N=[\epsilon^{-1}], \qquad \delta=\frac{1}{N+1}$$

and J the interval $\delta\leq u\leq 1+\delta$. We consider all pairs of integers q, r with

$$1\leq r\leq q\leq N, \qquad (q,r)=1 \tag{30}$$

and denote by J_{qr} the interval

$$|u-r/q| \leq \frac{1}{2qN}, \tag{31}$$

and by J_0 the set of all points of J not belonging to any J_{qr}.

LEMMA 8. *The intervals J_{qr} are contained in J and do not overlap each other. For any point u of the set J_0, a pair of integers q, r satisfying* (30) *exists, such that*

$$\frac{1}{2qN} < |u-r/q| < \frac{1}{qN}. \tag{32}$$

Proof. Let u be a point of J_{qr}. By (30) and (31)

$$u\geq r/q-\frac{1}{2qN}\geq 1/q\Big(1-\frac{1}{2N}\Big)\geq 1/N\Big(1-\frac{1}{N+1}\Big)=\delta,$$

$$u\leq r/q+\frac{1}{2qN}\leq 1+\frac{1}{2N}\leq 1+\delta;$$

hence u belongs to the interval J. If another interval $J_{q'r'}$ had an inner point u in common with J_{qr}, the inequality

$$| u - r'/q' | < \frac{1}{2q'N}$$

and (30), (31) lead to the contradiction

$$\frac{1}{qq'} \leq \frac{| rq' - qr' |}{qq'} = \left| \frac{r}{q} - \frac{r'}{q'} \right| < \frac{1}{2qN} + \frac{1}{2q'N} = \frac{q + q'}{2qq'N} \leq \frac{1}{qq'}.$$

Let now u be a point of J_0. Applying the theory of continued fractions we find a pair of integers q', r' such that

$$1 \leq q' \leq N, \qquad (q', r') = 1, \qquad | u - r'/q' | < \frac{1}{q'N}.$$

If $r' \leq 0$, we have

$$\frac{1}{N+1} = \delta \leq u < \frac{1}{q'N} \leq \frac{1}{N}$$

$$\left| u - \frac{1}{N} \right| < \frac{1}{N^2};$$

if $r' > q'$, we have

$$1 - \frac{1}{N} \leq 1 - \frac{1}{q'N} < u \leq 1 + \delta = 1 + \frac{1}{N+1}$$

$$\left| u - \frac{1}{1} \right| < \frac{1}{N};$$

hence the inequality

$$\left| u - \frac{r}{q} \right| < \frac{1}{qN}$$

holds also for at least one pair q, r satisfying (30). On the other hand, u not being a point of J_{qr},

$$\frac{1}{2qN} < \left| u - \frac{r}{q} \right|.$$

This completes the proof of the lemma.

Putting

$$S = \sum_{k=1}^{m} a_k x_k^2, \qquad \mathfrak{P} = \sum_{k=1}^{m} | a_k | x_k^2$$

with integral coefficients $a_k \neq 0$, we define the function

(33) $$f(u) = \sum_{x} \exp(-\pi\epsilon\mathfrak{P} + 2\pi i u S),$$

where the sum is extended over all integral values of the variables $x_1, \cdots, x_m$. Obviously

$$(34) \qquad f(u) = \prod_{k=1}^{m} f_k(u)$$

with

$$f_k(u) = \sum_{l=-\infty}^{\infty} \exp\{-\pi \mid a_k \mid (\epsilon \mp 2iu)l^2\} \qquad (k = 1, \cdots, m),$$

where the sign is defined by

$$\pm 1 = \frac{a_k}{\mid a_k \mid}.$$

Using the substitution $u = v + r/q$, we obtain

$$f_k(u) = \sum_{h=1}^{q} \rho_q^{ra_kh^2} \sum_{l=-\infty}^{\infty} \exp\{-\pi \mid a_k \mid (\epsilon \mp 2iv)(ql + h)^2\}.$$

Applying the well-known formula

$$\sum_{l=-\infty}^{\infty} \exp\{-\pi z(l + w)^2\} = z^{-\frac{1}{2}} \sum_{l=-\infty}^{\infty} \exp\{-(\pi/z)l^2 + 2\pi ilw\}$$

from the theory of theta functions with

$$z = z_k = \mid a_k \mid q^2(\epsilon \mp 2iv), \qquad w = h/q$$

and introducing the abbreviation

$$(35) \qquad \gamma_{kl}(r/q) = q^{-1} \sum_{h=1}^{q} \rho_q^{ra_kh^2 + lh},$$

we find

$$(36) \qquad f_k(u) = \mid a_k \mid^{-\frac{1}{2}} (\epsilon \mp 2iv)^{-\frac{1}{2}} \sum_{l=-\infty}^{\infty} \gamma_{kl}(r/q) \exp\{-(\pi/z_k)l^2\}.$$

Let now u be a point of J. By Lemma 8, a fraction r/q satisfying (30) exists such that $\mid v \mid = \mid u - r/q \mid < 1/qN$. The real part of

$$z_k^{-1} = \mid a_k \mid^{-1} q^{-2} \frac{\epsilon \pm 2iv}{\epsilon^2 + 4v^2}$$

has the value

$$\xi_k = \mid a_k \mid^{-1} (q^2\epsilon + 4q^2v^2\epsilon^{-1})^{-1}.$$

Since $q \leq N$ and $N \leq \epsilon^{-\frac{1}{2}} < N + 1$, we have

$$q^2\epsilon + 4q^2v^2\epsilon^{-1} \leq 1 + 4(1 + N^{-1})^2 \leq 17$$

and therefore

$$(37) \qquad \mid \exp(-\pi/z_k) \mid = \exp(-\pi\xi_k) \leq \exp(-\pi \mid a_k \mid^{-1}/17) \qquad (k = 1, \cdots, m).$$

In the following estimates a formula of the type $A = O(\epsilon^c)$ with real exponent c means that the inequality $\mid A \mid < K\epsilon^c$ holds with a certain constant

K depending only upon the given quadratic form S and not upon ϵ or any other parameter involved in A.

LEMMA 9. *If u is any point of J_0, then*

$$f(u) = O(\epsilon^{-m/4}).$$

Proof. By Lemma 1 and (35)

$$(38) \qquad | \gamma_{kl}(r/q) | \leq q^{-\frac{1}{2}}(2a_k, q)^{\frac{1}{2}} \leq | 2a_k |^{\frac{1}{2}} q^{-\frac{1}{2}};$$

by Lemma 8

$$(39) \qquad | \epsilon \mp 2iv | \geq | 2v | = 2 | u - r/q | > 1/qN \geq q^{-1}\epsilon^{\frac{1}{2}}.$$

From (36), (37), (38), (39) we infer

$$f_k(u) = O(\epsilon^{-\frac{1}{4}})$$

and consequently by (34) the assertion of our lemma.

Let n of the numbers $a_1, \cdots, a_m$ be positive and $m - n$ negative. We define

$$F(v) = (\epsilon - 2iv)^{-n/2}(\epsilon + 2iv)^{-(m-n/2)}$$

and use the symbols D and $G_q(r)$ in their former meaning.

LEMMA 10. *If u is any point of J_{qr}, then*

$$f(u) = | D |^{-\frac{1}{2}} q^{-m} G_q(r) F(v) + O(\epsilon^{-m/4}).$$

Proof. By (36), (37), (38)

$$f_k(u) = | a_k |^{-\frac{1}{2}}(\epsilon \mp 2iv)^{-\frac{1}{2}} \{\gamma_{k0}(r/q) + q^{-\frac{1}{2}} \exp(-\pi\xi_k) O(1)\};$$

hence by (3) and (34)

$$f(u) = | D |^{-\frac{1}{2}} q^{-m} G_q(r) F(v) + q^{-(m/2)}(\epsilon^2 + 4v^2)^{-(m/4)}$$
$$\{-1 + \prod_{k=1}^{m} (1 + \exp[-\pi\xi_k])\} O(1).$$

Now the assertion follows from the estimate

$$q^{-(m/2)}(\epsilon^2 + 4v^2)^{-(m/4)} \exp(-\pi\xi_k) = | a_k \xi_k \epsilon^{-1} |^{m/4} \exp(-\pi\xi_k) = O(\epsilon^{-m/4}).$$

LEMMA 11. *If $m > 2$ and $0 < n < m$, then*

$$(40) \quad \int_{-1/2qN}^{1/2qN} F(v)\,dv = \frac{\pi\Gamma(m/2 - 1)}{\Gamma(n/2)\Gamma([m-n]/2)} (2\epsilon)^{1-m/2} + q^{m/2-1} O(\epsilon^{1/2-m/4}).$$

Proof. Let W be the integral on the left-hand side of (40) and W_0 the same integral between the limits $-\infty$ and ∞. Obviously

$$(41)\qquad | W - W_0 | < 2 \int_{1/2qN}^{\infty} (2v)^{-m/2} dv = \frac{(qN)^{m/2-1}}{m/2-1} = q^{m/2-1} O(\epsilon^{1/2-m/4}),$$

$$W_0 = \tfrac{1}{2}\epsilon^{1-m/2} \int_{-\infty}^{\infty} (1-iv)^{-n/2}(1+iv)^{-(m-n)/2} dv.$$

Since

$$\Gamma(n/2)(1-iv)^{-n/2} = \int_0^{\infty} x^{n/2-1} \exp\{-(1-iv)x\} dx$$

and

$$\int_{-\infty}^{\infty} \exp(ivx)(1+iv)^{-(m-n)/2} dv = \frac{2\pi}{\Gamma([m-n]/2)} x^{(m-n)/2-1} e^{-x},$$

we obtain

$$(42)\qquad W_0 = \frac{\pi\epsilon^{1-m/2}}{\Gamma(n/2)\Gamma([m-n]/2)} \int_0^{\infty} x^{m/2-2} \exp(-2x) dx$$

$$= \frac{\pi\Gamma(m/2-1)}{\Gamma(n/2)\Gamma([m-n]/2)} (2\epsilon)^{1-m/2}.$$

By (41) and (42), the lemma follows.

3. Proof of Theorems 3 and 4. Let S be indefinite. By (2) and (33)

$$A(\epsilon) = \int_J f(u) du;$$

hence by Lemma 8

$$(43)\qquad A(\epsilon) = \int_{J_0} f(u) du + \sum_{q,r} \int_{J_{qr}} f(u) du,$$

where q and r run over all integers satisfying (30). By Lemma 9

$$(44)\qquad \int_{J_0} f(u) du = O(\epsilon^{-m/4});$$

by (31) and Lemma 10

$$\int_{J_{qr}} f(u) du = | D |^{-\frac{1}{2}} q^{-m} G_q(r) \int_{-1/2qN}^{1/2qN} F(v) dv + 1/qN\, O(\epsilon^{-m/4}).$$

Using the abbreviation

$$(45)\qquad \omega = | D |^{-\frac{1}{2}} \frac{\pi\Gamma(m/2-1)}{\Gamma(n/2)\Gamma([m-n]/2)} 2^{1-m/2}$$

we obtain by (5) and Lemma 11

$$\int_{J_{qr}} f(u)\,du = \omega q^{-m} G_q(r) \epsilon^{1-m/2} + 1/qN\, O(\epsilon^{-m/4}). \tag{46}$$

By (6), (30), (43), (44), (46) we find

$$A(\epsilon) = \omega \epsilon^{1-m/2} \sum_{q=1}^{N} H_q + O(\epsilon^{-m/4}). \tag{47}$$

If $m > 4$, then $1 - m/2 < - m/4$ and by (45), (47) and Lemma 4

$$\lim_{\epsilon \to 0} \epsilon^{m/2-1} A(\epsilon) = \omega \sum_{q=1}^{\infty} H_q = \omega\sigma > 0.$$

This proves Theorem 3.

If $m = 4$ and D is a square and S a zero-form in all R_p, the expression $(\log N)^{-1} \sum_{q=1}^{N} H_q$ tends by Lemma 7 for $N \to \infty$ to a positive limit σ_0. Since $\epsilon^{-\frac{1}{2}} - 1 < N \leq \epsilon^{-\frac{1}{2}}$, we obtain by (47) the relationship

$$\lim_{\epsilon \to 0} \frac{\epsilon}{\log \epsilon^{-1}} A(\epsilon) = \tfrac{1}{2}\omega\sigma_0 > 0,$$

and Theorem 4 is proved.

4. Proof of Theorem 1.

LEMMA 12. *Let the quadratic form S be a zero-form in R and c any rational number. Then the equation $S = c$ has a solution in rational numbers.*

Proof. Without loss of generality we may assume

$$S(x) = \sum_{k=1}^{m} a_k x_k^2, \qquad \sum_{k=1}^{m} a_k r_k^2 = 0$$

with rational numbers a_k, r_k and $a_1 r_1 \neq 0$. Putting

$$\rho = c(4a_1 r_1^2)^{-1}, \qquad x_1 = (1+\rho) r_1, \qquad x_k = (1-\rho) r_k \qquad (k = 2, \cdots, m),$$

we obtain $S(x) = 4\rho a_1 r_1^2 = c$, and the lemma is proved.

Let now S be a quadratic form with coefficients in P. A matrix $\mathfrak{C}$ with elements c_{kl} $(k, l = 1, \cdots, m)$ of P is called a *unit of S in* P, if the linear substitution

$$x_k \to \sum_{l=1}^{m} c_{kl} x_l \qquad (k = 1, \cdots, m)$$

carries S into itself.

LEMMA 13. *Let P be one of the fields R, R_p, R_∞ and $S(x)$ a diagonal*

form with coefficients in P. *If* $x_1, \cdots, x_m$ *are any given numbers of* P, *not all zero, then there exists in* P *a unit* $\mathfrak{C}$ *of* S, *such that* x_1 *is carried into a number* $\neq 0$.

Proof. The assertion is trivial in the case $x_1 \neq 0$, and we may assume $x_1 = 0$, $x_h \neq 0$, where h is one of the numbers $2, \cdots, m$. Let λ be a parameter, $S = \sum_{k=1}^{m} a_k x_k^2$ and

$$\alpha = \frac{a_1 a_h - \lambda^2}{a_1 a_h + \lambda^2}, \qquad \beta = \frac{2\lambda}{a_1 a_h + \lambda^2}.$$

Since $\alpha^2 + a_1 a_h \beta^2 = 1$, the linear substitution

$$(48) \qquad x_1 \to \alpha x_1 + \beta a_h x_h, \qquad x_h \to -\beta a_1 x_1 + \alpha x_h, \qquad x_k \to x_k \quad (k \neq 1, h)$$

carries S into itself. Choosing for λ any number of P, such that $\lambda \neq 0$ and $\lambda^2 \neq -a_1 a_h$, we obtain from (48) a unit of S with the required property.

LEMMA 14.[8] *Let* P *be one of the fields* R, R_p R_∞ *and* S, T *two quadratic forms in the* m *variables* $x_1, \cdots, x_m$ *with coefficients in* P. *If the quadratic forms* $ax_0^2 + S$ *and* $ax_0^2 + T$ *in the* $m + 1$ *variables* $x_0, \cdots, x_m$ *are equivalent in* P, *then also* S *and* T *are equivalent in* P.

Proof. Obviously we may assume that $a = 1$ and that S is a diagonal form. Let $\mathfrak{C}$ be any unit of $x_0^2 + S$ in P. If $\mathfrak{G}$ is the matrix of a linear substitution

$$x_k \to \sum_{l=0}^{m} g_{kl} x_l \qquad (k = 0, \cdots, m)$$

carrying $x_0^2 + S$ into $x_0^2 + T$, with coefficients in P, then $\mathfrak{C}\mathfrak{G}$ has the same property. Since not all elements g_{k0} $(k = 0, \cdots, m)$ of the first column of $\mathfrak{G}$ are zero, we may assume, by Lemma 13, that g_{00} is a number $b \neq 0$. Denoting by $\mathfrak{c}$ and $\mathfrak{d}$ the columns of the coefficients g_{k0} $(k = 1, \cdots, m)$ and g_{0l} $(l = 1, \cdots, m)$, by $\mathfrak{H}$ the matrix of the m^2 elements g_{kl} $(k, l = 1, \cdots, m)$, by $\mathfrak{S}$, $\mathfrak{T}$ the matrices of S, T and using the abbreviation $\mathfrak{M}'\mathfrak{S}\mathfrak{M} = \mathfrak{S}[\mathfrak{M}]$, we obtain the matrix equations

$$b^2 + \mathfrak{S}[\mathfrak{c}] = 1, \qquad \mathfrak{d}b + \mathfrak{H}'\mathfrak{S}\mathfrak{c} = 0, \qquad \mathfrak{d}\mathfrak{d}' + \mathfrak{S}[\mathfrak{H}] = \mathfrak{T}.$$

Hence $\mathfrak{d} = -\mathfrak{H}'\mathfrak{S}\mathfrak{c}b^{-1}$ and

$$(\mathfrak{S} + b^{-2}\mathfrak{S}\mathfrak{c}\mathfrak{S})[\mathfrak{H}] = \mathfrak{T}.$$

Defining

$$\mathfrak{F} = \mathfrak{E} + (b^2 \pm b)^{-1}\mathfrak{c}\mathfrak{c}'\mathfrak{S}, \qquad \mathfrak{W} = \mathfrak{S}\mathfrak{c}\mathfrak{c}'\mathfrak{S},$$

---

[8] E. Witt, "Theorie der quadratischen Formen in beliebigen Körpern," *Journal für die reine und angewandte Mathematik*, vol. 176 (1937), pp. 31-44.

where $\mathfrak{E}$ is the ordinary unit matrix and the sign is determined such that $b^2 \pm b \neq 0$, we obtain

$$\mathfrak{S}[\mathfrak{F}] = \mathfrak{S} + 2(b^2 \pm b)^{-1}\mathfrak{W} + (b^2 \pm b)^{-2}(1 - b^2)\mathfrak{W} = \mathfrak{S} + b^{-2}\mathfrak{W}$$
$$\mathfrak{S}[\mathfrak{F}\mathfrak{H}] = \mathfrak{T},$$

and the lemma is proved.

We come now to the proof of Theorem 1. We consider two quadratic forms

$$S = \sum_{k=1}^{m} a_k x_k^2, \qquad T = \sum_{k=1}^{m} b_k y_k^2$$

with integral coefficients $a_k, b_k \neq 0$, which are equivalent in all R_p and R_∞. Denoting the determinants of S and T by D_1 and D_2, we infer that the product $D_1 D_2$ is a square number in all R_p and in R_∞, hence the square of a rational integer. This proves the theorem in the case $m = 1$. Let us assume that $m > 1$ and that the theorem holds for $m - 1$ instead of m.

Consider now the quadratic form $V = S - T$ in the $2m$ variables x_k, y_k $(k = 1, \cdots, m)$. Its determinant is $(-1)^m D_1 D_2$ and hence is a square for any even m and in particular for $m = 2$. On account of the equivalence of S and T in all R_p and R_∞, we see that V is a zero-form in all these fields. Therefore we may apply the Theorems 3 and 4 to the quadratic form V. Writing

$$W = \sum_{k=1}^{m} (|\,a_k\,|\, x_k^2 + |\,b_k\,|\, y_k^2), \qquad B(\epsilon) = \sum_{V=0} \exp(-\pi\epsilon W) \qquad (\epsilon > 0),$$

where the summation extends over all integral solutions of $V = 0$, we conclude from those theorems that the expression $\epsilon^{m-1}B(\epsilon)$ tends for $\epsilon \to 0$ to a positive limit, if $m > 2$, and in the case $m = 2$ the same is true for the expression $\dfrac{\epsilon}{\log \epsilon^{-1}} B(\epsilon)$.

The solutions of $V = 0$ are identical with the solutions of $S = T$. Let us suppose, for a moment, that for all these solutions either every $x_k = 0$ $(k = 1, \cdots, m)$ or every $y_k = 0$ $(k = 1, \cdots, m)$. Then certainly

$$B(\epsilon) \leq \{ \sum_{l=-\infty}^{\infty} \exp(-\pi\epsilon l^2) \}^m,$$

and consequently $\epsilon^{m/2}B(\epsilon)$ is bounded for $\epsilon \to 0$, in contradiction to the asymptotic behavior of $B(\epsilon)$.

Hence there exists an integral solution $x = x^{(1)}$, $y = y^{(1)}$ of $S(x) = T(y)$ with $x^{(1)} \neq 0$, $y^{(1)} \neq 0$. If $S(x^{(1)}) = T(y^{(1)}) = 0$, then S and T are zero-forms in R, and we can construct, by Lemma 12, an integral solution $x^{(2)}$, $y^{(2)}$ of $S(x) = T(y) \neq 0$. Hence we may assume, in any case, that $S(x^{(1)}) =$

$T(y^{(1)}) = a \neq 0$. Now we can find in R two matrices $\mathfrak{A}$ and $\mathfrak{B}$ with non-vanishing determinants and the first columns $x^{(1)}$ and $y^{(1)}$. Performing on S and T linear transformations with these matrices, we obtain two quadratic forms S_1 and T_1 with the same coefficient a of the square of the first variable, and these quadratic forms are again equivalent in all R_p and R_∞. Moreover, by completing squares, we may assume that $S_1 - ax_1{}^2 = S_2$ and $T_1 - ay_1{}^2 = T_2$ depend only upon the last $m-1$ variables. By Lemma 14, S_2 and T_2 are equivalent in all R_p and R_∞. Since our theorem holds for $m-1$ variables, S_2 and T_2 are equivalent in R, hence also S_1 and T_1 and finally S and T.

5. Proof of Theorem 2. We denote by $\mathfrak{R}^{(1)}, \cdots, \mathfrak{R}^{(h)}$ the different diagonal matrices having the m diagonal elements ± 1; their number is $h = 2^m$.

LEMMA 15. *Let $\mathfrak{L}$ be a matrix in* P *with non-vanishing determinant. There exists at least one matrix $\mathfrak{R}^{(g)}$ such that the determinant* $|\mathfrak{L} - \mathfrak{R}^{(g)}| \neq 0$.

Proof. Let $\mathfrak{D}$ be the diagonal matrix with indeterminate diagonal elements $\lambda_1, \cdots, \lambda_m$ and take $\mathfrak{D}_l = \mathfrak{R}^{(l)}\mathfrak{D}$ $(l = 1, \cdots, h)$. The determinant $|\mathfrak{L} - \mathfrak{D}|$ is a linear function of any single λ_k $(k = 1, \cdots, m)$ and the same holds for the sum

$$L = \sum_{l=1}^{h} |\mathfrak{L} - \mathfrak{D}_l|.$$

Since the matrices $\mathfrak{R}^{(l)}$ form a group under multiplication, the function L is not changed, if $\mathfrak{D}$ is replaced by $\mathfrak{R}\mathfrak{D}$, when $\mathfrak{R}$ denotes any of the matrices $\mathfrak{R}^{(l)}$. Consequently L is an even function of any single variable λ_k $(k = 1, \cdots, m)$. This proves that L is a constant. Taking in particular $\mathfrak{D} = \mathfrak{E}$ and $\mathfrak{D} = 0$, we obtain

$$\sum_{l=1}^{h} |\mathfrak{L} - \mathfrak{R}^{(l)}| = h\,|\mathfrak{L}| \neq 0.$$

LEMMA 16. *Let S be a quadratic form with the matrix $\mathfrak{S}$ and with coefficients in* P, *where* P *is one of the fields R, R_p, R_∞. If $\mathfrak{A}$ denotes any skew-symmetric matrix in* P *such that the determinant* $|\mathfrak{A} + \mathfrak{S}| \neq 0$, *then*

$$\mathfrak{C} = (\mathfrak{A} + \mathfrak{S})^{-1}(\mathfrak{A} - \mathfrak{S}) \tag{49}$$

is a unit of S in P *and* $|\mathfrak{E} - \mathfrak{C}| \neq 0$. *Vice versa, for any unit $\mathfrak{C}$ of S in* P *with* $|\mathfrak{E} - \mathfrak{C}| \neq 0$ *a skew-symmetric matrix $\mathfrak{A}$ in* P *exists, such that* $|\mathfrak{A} + \mathfrak{S}| \neq 0$ *and* (49) *holds.*

Proof. The units $\mathfrak{C}$ of S are the solutions of the matrix equation

$$\mathfrak{S}[\mathfrak{C}] = \mathfrak{S}. \tag{50}$$

If moreover $|\mathfrak{E}-\mathfrak{C}|\neq 0$, we define $\mathfrak{M}=2(\mathfrak{E}-\mathfrak{C})^{-1}$ and

$$\mathfrak{A}=\mathfrak{S}\mathfrak{M}-\mathfrak{S}. \tag{51}$$

Then $|\mathfrak{M}|\neq 0$, $|\mathfrak{A}+\mathfrak{S}|\neq 0$ and

$$\mathfrak{C}\mathfrak{M}=\mathfrak{M}-2\mathfrak{E}. \tag{52}$$

By (51) and (52), the formula (49) follows. By (50) and (52)

$$0=\mathfrak{S}[\mathfrak{M}]-\mathfrak{S}[\mathfrak{M}-2\mathfrak{E}]=2(\mathfrak{S}\mathfrak{M}+\mathfrak{M}'\mathfrak{S}-2\mathfrak{S})=2(\mathfrak{A}+\mathfrak{A}'); \tag{53}$$

hence $\mathfrak{A}$ is skew-symmetric. On the other hand, if $\mathfrak{A}$ is any skew-symmetric matrix in P and $|\mathfrak{A}+\mathfrak{S}|\neq 0$, we define $\mathfrak{M}$ and $\mathfrak{C}$ by (51) and (52). Then $(\mathfrak{E}-\mathfrak{C})\mathfrak{M}=2\mathfrak{E}$, $|\mathfrak{M}|\neq 0$, $|\mathfrak{E}-\mathfrak{C}|\neq 0$ and (50) follows from (52) and (53).

LEMMA 17. *Let S be a quadratic form with integral p-adic coefficients and $p\neq 2$. Then S is equivalent in J_p to a diagonal form.*

Proof. The assertion is trivial for $m=1$ and we may apply induction with respect to m. Let p^a be the largest power of p dividing all coefficients s_{kl} $(k,l=1,\cdots,m)$ of S. If all diagonal elements are divisible by p^{a+1}, then a certain coefficient s_{ab} $(a\neq b)$ is exactly divisible by p^a. By the substitution $x_b\to x_a+x_b$, $x_k\to x_k$ $(k\neq b)$, the quadratic form S is replaced by the equivalent form $S+s_{bb}x_a^2+2x_a\sum_{l=1}^{m}s_{bl}x_l$, and now x_a^2 has the coefficient $s_{aa}+2s_{ab}+s_{bb}$ divisible exactly by the power p^a, whereas all other coefficients are divisible at least by this power. Hence we may assume that already s_{aa} is not divisible by p^{a+1}. Then the difference

$$S-s_{aa}(x_a+\sum_{k\neq a}\frac{s_{ak}}{s_{aa}}x_k)^2$$

is a quadratic form with integral p-adic coefficients and only $m-1$ variables x_k $(k\neq a)$. This proves the lemma.

The proof of Theorem 2 proceeds in the following way. Let $\mathfrak{S}$ and $\mathfrak{T}$ be the matrices of two quadratic forms S and T with integral coefficients, which are equivalent in all J_p and J_∞. For any prime number p, a matrix $\mathfrak{B}_p$ of J_p exists such that $\mathfrak{S}[\mathfrak{B}_p]=\mathfrak{T}$. Moreover, by Theorem 1, the equation $\mathfrak{S}[\mathfrak{B}]=\mathfrak{T}$ holds for a certain matrix $\mathfrak{B}$ of R.

We determine a matrix $\mathfrak{W}$ of R, such that $\mathfrak{S}[\mathfrak{W}]=\mathfrak{S}_2$ becomes a diagonal matrix. Applying Lemma 15 for $\mathrm{P}=R_2$ and $\mathfrak{L}_2=\mathfrak{W}^{-1}\mathfrak{B}_2\mathfrak{B}^{-1}\mathfrak{W}$, we find a certain diagonal matrix $\mathfrak{K}_2$ with the diagonal elements ± 1 and $|\mathfrak{L}_2-\mathfrak{K}_2|\neq 0$.

Obviously $\mathfrak{S}_2[\mathfrak{K}_2] = \mathfrak{S}_2$. The matrix $\mathfrak{B}_0 = \mathfrak{W}\mathfrak{K}_2\mathfrak{W}^{-1}\mathfrak{B}$ has rational elements and satisfies

$$\mathfrak{S}[\mathfrak{B}_0] = \mathfrak{T}, \qquad \mathfrak{B}_2 - \mathfrak{B}_0 = \mathfrak{W}(\mathfrak{L}_2 - \mathfrak{K}_2)\mathfrak{W}^{-1}\mathfrak{B}, \qquad |\,\mathfrak{B}_0 - \mathfrak{B}_2\,| \neq 0.$$

Let now p be any odd prime number. By Lemma 17 a matrix $\mathfrak{W}_p$ exists in J_p such that $\mathfrak{S}[\mathfrak{W}_p] = \mathfrak{S}_p$ is a diagonal matrix and also $\mathfrak{W}_p^{-1}$ belongs to J_p. Applying Lemma 15 for $\mathrm{P} = R_p$ and $\mathfrak{L}_p = \mathfrak{W}_p^{-1}\mathfrak{B}_0\mathfrak{B}_p^{-1}\mathfrak{W}_p$, we obtain a diagonal matrix $\mathfrak{K}_p$ with the diagonal elements ± 1 and $|\,\mathfrak{L}_p - \mathfrak{K}_p\,| \neq 0$. Again $\mathfrak{S}_p[\mathfrak{K}_p] = \mathfrak{S}_p$. Now the matrix $\mathfrak{B}_p{}^* = \mathfrak{W}_p\mathfrak{K}_p\mathfrak{W}_p^{-1}\mathfrak{B}_p$ belongs to J_p and satisfies

$$\mathfrak{S}[\mathfrak{B}_p{}^*] = \mathfrak{T}, \qquad \mathfrak{B}_0 - \mathfrak{B}_p{}^* = \mathfrak{W}_p(\mathfrak{L}_p - \mathfrak{K}_p)\mathfrak{W}_p^{-1}\mathfrak{B}_p, \qquad |\,\mathfrak{B}_0 - \mathfrak{B}_p{}^*\,| \neq 0.$$

In the case $p = 2$, we define $\mathfrak{B}_p{}^* = \mathfrak{B}_p$. Then we have found a matrix $\mathfrak{B}_0$ in R and a matrix $\mathfrak{B}_p{}^*$ in every J_p with the properties

$$\mathfrak{S}[\mathfrak{B}_0] = \mathfrak{S}[\mathfrak{B}_p{}^*] = \mathfrak{T}, \qquad |\,\mathfrak{B}_0{}^* - \mathfrak{B}_p{}^*\,| \neq 0.$$

Let now q be any given positive integer and p a prime factor of q. We use Lemma 16 for $\mathrm{P} = R_p$ and the unit $\mathfrak{C}_p = \mathfrak{B}_p{}^*\mathfrak{B}_0{}^{-1}$ of S in R_p. Since the condition $|\,\mathfrak{E} - \mathfrak{C}_p\,| \neq 0$ is fulfilled, a skew-symmetric matrix $\mathfrak{A}_p$ exists in R_p such that $|\,\mathfrak{A}_p + \mathfrak{S}\,| \neq 0$ and

$$\mathfrak{B}_p{}^* = (\mathfrak{A}_p + \mathfrak{S})^{-1}(\mathfrak{A}_p - \mathfrak{S})\mathfrak{B}_0.$$

If β is an arbitrarily large integer, a skew-symmetric matrix $\mathfrak{A}$ with rational elements exists which satisfies the congruences $\mathfrak{A} \equiv \mathfrak{A}_p \pmod{p^\beta}$ for all prime factors p of q. Now $|\,\mathfrak{A}_p + \mathfrak{S}\,| \neq 0$ and moreover all elements of $\mathfrak{B}_p{}^*$ are p-adic integers. Hence for sufficiently large β, the inequality $|\,\mathfrak{A} + \mathfrak{S}\,| \neq 0$ holds and the rational matrix

$$\mathfrak{B}^* = (\mathfrak{A} + \mathfrak{S})^{-1}(\mathfrak{A} - \mathfrak{S})\mathfrak{B}_0$$

is p-adically integral, for all prime factors p of q. This means that the denominators of the elements of $\mathfrak{B}^*$ are relative-prime to q. On the other hand, by Lemma 16,

$$\mathfrak{S}[\mathfrak{C}^*] = \mathfrak{S}[\mathfrak{B}_0] = \mathfrak{T}.$$

Hence S represents T rationally without essential denominator. In this result we may interchange S and T. Consequently S and T are semiequivalent and the proof of Theorem 2 is complete.

THE INSTITUTE FOR ADVANCED STUDY.